

OM of: US-09-277-074-10 to: GenEmbl:\* out\_format : pfs

Date: Nov 15, 2000 10:16 PM

About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=frame+g2n.model -DEV=xlp  
-Q/Cgn2\_1/USPTO\_SPOOL/US9277074/runat\_14112000\_120308\_22616/app\_query.fasta\_1.67  
-DB=GenEmbl -QFMT=fastcap -SUFFIX=rge -GAPOP=12.000 -GAPEXT=4.000  
-MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=4.500  
-GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000  
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000  
-DELEXT=7.000 -START=1 -MATRIX=blomsum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0  
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0  
-MAXLEN=2000000000 -USER=US09277074\_CGNI\_1\_3727 -NCPU=6  
-TCPU=3 -LONGLOG -NO\_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-277-074-10

Query length: 9

Database: GenEmbl.\*

Database sequences: 1033670

Database length: -2111177393

Search time (sec): 1109.000000

score\_list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
gb_pat:AR034479	+	43.00	11.26	3768	AR034479 Sequence 1 from patent
gb_om:AB008451	+	43.00	139.97	3780	AB008451 Canis familiaris mRNA
gb_pat:121129	+	43.00	139.62	3955	121129 Sequence 14 from patent
gb_pat:159750	+	43.00	139.62	3955	159750 Sequence 14 from patent
gb_om:RNNRUR	+	43.00	139.62	3955	X03362 Rat mRNA for neu oncogene
gb_om:HANEU	+	43.00	139.41	4062	D16295 Syrian golden hamster
gb_pr6:HSRBR2R	+	43.00	138.65	13.38	X03363 Human c-erbB-2 mRNA
gb_pat:121124	+	43.00	138.55	4530	121124 Sequence 9 from patent
gb_pat:159745	+	43.00	138.55	4530	159745 Sequence 9 from patent
gb_pr7:HUMHER2A	+	43.00	138.55	4530	M11730 Human tyrosine kinase
gb_v12:RRNASEG3	-	43.00	110.87	472.06	154278
gb_v12:RRNASEG3	-	43.00	136.57	17.47	X82047 Reovirus sp. 1.318kb RN
gb_pr1:AC004998	-	40.00	100.21	1.9e+03	AC004998 Homo sapiens PAC cl
gb_htg21:AL359545	+	39.00	94.99	3.8e+03	AL359545 Homo sapiens chromo
gb_htg19:AL162394	+	39.00	94.10	4.1e+03	AL162394 Homo sapiens chromo
gb_pr2:AC006127	+	38.00	102.66	1.4e+03	AC006127 Homo sapiens chromo
gb_htg10:AC023733	+	38.00	100.76	1.7e+03	AC023733 Drosophila melanoga
gb_htg9:AC019575	+	38.00	97.34	2.7e+03	AC019575 Drosophila melanoga
em_in:DM25P	-	38.00	96.41	3.0e+03	AL49405 Drosophila melanogaste
gb_htg7:AC019670	+	38.00	95.00	3.6e+03	AC019670 Drosophila melanoga
gb_htg23:HSJ988112	+	38.00	92.46	5.0e+03	AC010991 Homo sapiens clone
gb_htg17:AC073721	+	38.00	80.37	6.6e+03	AL121837 Homo sapiens chromo
gb_in1:AE003641	+	38.00	86.21	1.1e+04	AE003641 Drosophila melanoga
gb_in3:DR05ADH01	+	38.00	86.10	1.1e+04	AE003407 Drosophila melanoga
gb_in3:DR05ADH01	+	38.00	86.02	1.1e+04	AE003476 Drosophila melanoga
gb_htg6:AC018041	+	37.00	101.34	1.6e+03	AC018041 Drosophila melanoga
gb_htg8:AC020614	+	37.00	97.73	2.8e+03	AC020614 Homo sapiens chromo
gb_p12:F2P3	+	37.00	94.39	3.9e+03	AC020003 Homo sapiens chromo
gb_p12:F2P3	+	37.00	93.25	4.5e+03	AF080120 Arabidopsis thaliana
gb_htg11:AC025257	-	37.00	88.16	8.7e+03	AL049876 Arabidopsis thaliana
gb_htg17:AC074274	-	37.00	87.67	9.3e+03	AC074274 Homo sapiens chromo
gb_htg6:AC016458	+	37.00	87.43	9.5e+03	AC016458 Homo sapiens clone
gb_htg8:AC021660	+	37.00	87.09	1.0e+04	AC021660 Homo sapiens chromo
gb_htg16:AC069362	-	37.00	86.71	1.0e+04	AC069362 Homo sapiens clone
gb_htg5:AC013632	-	37.00	86.53	1.1e+04	AC013632 Homo sapiens clone

gb\_htg15:AC068156 + 37.00 86.48 1.1e+04 176319 ! AC068156 Homo sapiens chr  
gb\_htg22:AP002501 + 37.00 86.33 1.1e+04 179844 ! AP002501 Homo sapiens chr  
gb\_htg3:AC010892 + 37.00 86.15 1.1e+04 184000 ! AC010892 Homo sapiens chr  
gb\_htg16:AC068986 + 37.00 86.12 1.1e+04 184561 ! AC068986 Homo sapiens chr  
seq\_name: gb\_pat:AR034479

seq\_documentation\_block:  
LOCUS AR034479 3768 bp DNA 29-SEP-1999  
DEFINITION Sequence 1 from patent US 5869445.  
ACCESSION AR034479  
VERSION AR034479.1 GI:5950084  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 3768)  
AUTHORS Cheever, M.A. and Disis, M.L.  
TITLE Methods for eliciting or enhancing reactivity to HER-2/neu protein  
JOURNAL Patent: US 5869445-A 1 09-FEB-1999;  
FEATURES  
Location/Qualifiers  
source 1..3768  
BASE COUNT 759 a 1171 c 1119 g 719 t  
ORIGIN

alignment\_scores:  
Quality: 43.00 Length: 9  
Ratio: 4.778 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-277-074-10 x AR034479 ..

Align seg 1/1 to: AR034479 from: 1 to: 3768

1 LysilePheGlySerLeuAlaPheLeu 9  
|||||  
1105 AGATCTTTGGGAGCTGGCATTTCTG 1131

seq\_name: gb\_om:AB008451

seq\_documentation\_block:  
LOCUS AB008451 3780 bp mRNA 30-OCT-1997  
DEFINITION Canis familiaris mRNA for erbB-2, complete cds.  
ACCESSION AB008451  
VERSION AB008451.1 GI:2575866  
KEYWORDS erbB-2.  
SOURCE Canis familiaris cDNA to mRNA.  
ORGANISM Canis familiaris  
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
Carnivora; Fissipedia; Canidae; Canis.  
REFERENCE 1 (bases 1 to 3780)  
AUTHORS Yokota, H.  
TITLE Direct Submission  
JOURNAL Submitted (23-OCT-1997) to the DDBJ/EMBL/GenBank databases. Hiroshi  
Yokota, Rakuno Gakuen University, Veterinary Biochemistry;  
Bunkyo-daimeidori-machi, Ebetsu, Hokkaido 069, Japan  
(E-mail: BXA03503@niftyserve.or.jp, Tel: 011-386-1111,  
Fax: 011-387-5890)

REFERENCE 2 (bases 1 to 3780)  
Yokota, H.  
CDNA cloning of erbB-2 from canine mammary gland  
Published Only in Database (1997) In press  
JOURNAL Location/Qualifiers  
FEATURES  
source 1..3780  
/organism="Canis familiaris"  
/db\_xref="taxon:9615"  
1..3780  
/codon\_start=1  
/product="erbB-2"  
/protein\_id="BAA23127.1"

/db\_xref="GI:2575867"  
/translation="MELAAACRWGLLLALLPSGAAGTQVCTGDMKLRLPASPETHLD  
MLRLHVGCGVGVGNELTYLPANASLFLDIOEVGVYLIASHSVQRQLPQLRLRV  
RGLQFLFDNALAVLNDGPLEGGIPAPGAAGGLRELOLRSLTEILKGGVLCQSPQ  
LCHQDITLMKDFVFKNNQLALITIDNRFACPPCPACKDAHCWAGSSDCQSLRT  
VCAGGACRCKGPOPTDCCHQCAAGCTGPKHSDCLACLHFNHSGIGELHCPALVTNT  
DTESWNPESGRVTFGASCYTSQPYAVLSTDVGSCITLVCPLNNQVTAEDGTORCEK  
SKPCARVCYGLGHEHLREVRVAVTSANIQFAGCKKIFGSLAFIPESFDGDPASNTAPL  
QPOLRVFEALEITGYLYISAWPDSLPNLSVFNLRVIRGRVLHDGAYSLTQGLGI  
SWGLSLRSLGSLALIHNRNLCFVHTVPMQLFRNPHQALLHSANRPEECVGE  
LACYPCAHCWGPPTQCVCNCSQFLRGQECVEECRVLOGLPREYVKDRYCLPCHSEC  
QPONGSVTCFSGEADQACAHYKDPPECVARGPSGVKPLDSEMPWKFADEGTGCP  
CPNCHSCADLDEKCPAERASPTVSIIAAVVGLLAVVGLVGLILKRRQKIR  
KYTMRLQLTELVEPLTPSGAMPNQAMRLKTELRLKVKVGLSGAFGVYKGIWIP  
DGENKIPVAIKVLRNTPSKANEILDEAYVMAGVSPVYVRLGICLSTVQVLTQ  
LMPYGLLHVHREHGRGLSGDOLLNMCVQAKGMSYLEDVRLVHRLAARNVLKSPN  
HVKITDFGLARLDIDTEYHAGGVPIKMALESIPPRFTHQSDVMSYVTVWEL  
MTFGAKGIPAREIPDLLEKGERLPQPICTIDVYMINVKWMDISECRPFRGLV  
AFPSRWARQPVVVIQNEEDLGPASPLDSTFYBSLLEDDMDGLVDAEEVLPQCGFF  
CPPTFGAGGTARRHRSSTRNGSGELTGLPSEEPKPSPLASEGAGSDVFDGD  
LGMGAAGLQSLSPQSPQRYSEDTVPLPETDKKVAPLTCSPOPEVYNQPEVWP  
OPPLALEGPLPSPAPAGATLERPKTLSPKTLSPGKNGVVDVFAFGSAVENPFLAPR  
GRAAPQHPPPAFSPADNLYYWDQPSERGSPPSTFEGTFTAEENPEYILGLDVPV"  
BASE COUNT 743 a 1162 c 1150 g 725 t  
ORIGIN

alignment\_scores:  
Quality: 43.00 Length: 9  
Ratio: 4.778 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-09-277-074-10 x AB008451 ..  
Align seg 1/1 to: AB008451 from: 1 to: 3780

1 LysilePheGlySerLeuAlaPheLeu 9  
|||||  
1105 AAGATCTTTGGAGCCCTGGCATTTTG 1131

seq\_name: gb\_pat:121129

seq\_documentation\_block:  
LOCUS 121129 3955 bp DNA PAT 07-OCT-1996

DEFINITION Sequence 14 from patent US 5518885.

ACCESSION 121129

VERSION 121129.1 GI:1601483

KEYWORDS

SOURCE

ORGANISM

Unclassified.

REFERENCE 1 (bases 1 to 3955)

AUTHORS Raziuddin;.. and Sarkar,F.H.

TITLE ERB2 promoter binding protein in neoplastic disease

JOURNAL Patent: US 5518885-A 14 21-MAY-1996;

FEATURES

Location/Qualifiers

source

1..3955

/organism="unknown"

BASE COUNT 842 a 1147 c 1136 g 830 t

ORIGIN

alignment\_scores:  
Quality: 43.00 Length: 9  
Ratio: 4.778 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-09-277-074-10 x I21129 ..  
Align seg 1/1 to: I21129 from: 1 to: 3955

1 LysilePheGlySerLeuAlaPheLeu 9  
|||||  
1133 AAGATCTTTGGAGCCCTGGCATTTTG 1159  
seq\_name: gb\_pat:159750

seq\_documentation\_block:

LOCUS 159750 3955 bp DNA PAT 07-OCT-1997

DEFINITION Sequence 14 from patent US 5654406.

ACCESSION 159750

VERSION 159750.1 GI:2478382

KEYWORDS

SOURCE

ORGANISM

Unclassified.

REFERENCE 1 (bases 1 to 3955)

AUTHORS Raziuddin;.. and Sarkar,F.Hoque.

TITLE Antibody to ERB2 promoter binding factor

JOURNAL Patent: US 5654406-A 14 05-AUG-1997;

FEATURES

Location/Qualifiers

source

1..3955

/organism="unknown"

BASE COUNT 842 a 1147 c 1136 g 830 t

ORIGIN

alignment\_scores:

Quality: 43.00 Length: 9

Ratio: 4.778 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-277-074-10 x I59750 ..

Align seg 1/1 to: I59750 from: 1 to: 3955

1 LysilePheGlySerLeuAlaPheLeu 9  
|||||  
1133 AAGATCTTTGGAGCCCTGGCATTTTG 1159

seq\_name: gb\_ro:RNNEUR

seq\_documentation\_block:

LOCUS RNNEUR 3955 bp mRNA ROD 30-MAR-1995

DEFINITION Rat mRNA for neu oncogene (p185) encoding an epidermal growth

factor receptor-related protein.

ACCESSION X03362

VERSION X03362.1 GI:56745

KEYWORDS

glycoprotein; kinase; neu oncogene; oncogene; transmembrane

protein; tyrosine kinase.

SOURCE

Norway rat.

ORGANISM

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;

Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE 1 (bases 1 to 3955)

AUTHORS Bargmann,C.I., Hung,M.C. and Weinberg,R.A.

TITLE The neu oncogene encodes an epidermal growth factor

receptor-related protein

JOURNAL Nature 319 (6050), 226-230 (1986)

MEDLINE 86118662

REFERENCE 2 (bases 1 to 3955)

AUTHORS Lofts,F.J., Hurst,H.C., Sternberg,M.J. and Gullick,W.J.

TITLE Specific short transmembrane sequences can inhibit transformation

by the mutant neu growth factor receptor in vitro and in vivo

JOURNAL Oncogene 8 (10), 2813-2820 (1993)

MEDLINE 93390956

FEATURES

Location/Qualifiers

source

1..3955

/organism="Rattus norvegicus"

/db\_xref="taxon:10116"

17..73

sig\_peptide

/product="put. signal peptide (aa -19 to -1)"

17..3799

CDS

```
/note="put. p185 precursor"
/codon_start=1
/protein_id="CAA27059.1"
/db_xref="GI:56746"
/db_xref="SWISS-PROT:P06494"
/translation="MIMELAAWCRWGLALLPPGIAGTQVCTGDMKRLRPASPET
HLDMLHYOCQVYVGNLEIYYPANASLSFLQDIOEVQYMLIAHVSQVRHVPLORLR
RIVRGTLQFEDKYALAVLNDPNDPOANASLPGRTPGRLQRLSLTEILKGGVLR
GNPOLCYDMVLMKDVFKNNQLAPVDIDTNRSRACPPACADKNHCWGESPEDCOI
LTGICISGCARCKGRFTDCHEQCACAGCTGPKHSDCLACHFNHSGICELHCPALV
TYNDTFESMNPGRYTFGASCTTCYNYLSTEVGSCITLCPNNOEVFAEDGTQR
CEKSPCARCYGELGMEHLRGARITSDNOEFGCKIFGSLAFLESFDPGSSG
IAPLRPEQLQVFEETLGYLISAWPDSRLDLSFQNLRIIRGLILHSGAYSILTLQ
GLGHSGLNSLRGLSGLAIHRNHLCFVHTVPDOLFRNPHQALLHSGNRPEDDL
CVSGLNLSLCACHCWGPCTOCVANCSEFLRGQECVEECRVKGLPREVVSQKRL
PCHPECPQNSERTCFGEADQACACAHYKSSCVACPCSPGVPDLSYFIWKPDE
EGICQPCINCHSCVDLDERGPAEQASPVFIATVEGVLFLILVYVVGILIR
RROKIRTYMRLLQETLVEPLTPSPGAMPQAQMRILKTELKRVKVGSGAFGTYY
KGIWTPDGENKIPVAKVLRNTPSKANKELDEAYVMAGVSPYVSRLLGICLTST
VOLVLMYPGKCLLDHVRHGRGLSQDLLNWCQIAGKMSYLEDVRLVHRDLAARNV
LVKSPNHVKITDFGLARLLDIDETEHADGCKVPLKWALESILRRBETHQDSVXG
VTVVELMTFGKPDYDIPARIPDLLEKGERLPQPTICTIDYIMIMKCMWIDSECRP
RFRLEVSFSRMARDPQRFVYQNEIDLGSPSPMDSTFYRSLLEDDDDMDGLVDABEYL
PQOGFSPDPTPGTSTAHRRSRSSSTRSGGELTGLPESEEGPPRSLAPSEGAGS
DVFGDLAMGYTKGLQSLSDHLSPLQRYSEDPTLPLPPTDGYVAPLACSPQPEYVN
QSEVOPOPPLTPEGLPVPVPAGATLERPKTLSPGKNGVVKDVAFGGAVERNPELVLP
REGTASPPHSPAPSAFDNLYYWDQNSSEGGPPSPNEFTPTAENPEYGLDVPV"
74..3796
/product="put. p185 (aa 1-1241)"
227..235
/note="pot. glycosylation site"
587..595
/note="pot. glycosylation site"
803..811
/note="pot. glycosylation site"
1619..1627
/note="pot. glycosylation site"
1742..1750
/note="pot. glycosylation site"
1916..1924
/note="pot. glycosylation site"
2087..2089
/note="pot. phosphorylation site"
2660..2662
/note="pot. phosphorylation site"
3446..3448
/note="pot. phosphorylation site"
3695..3697
/note="pot. phosphorylation site"
3723..3725
/note="pot. phosphorylation site"
BASE COUNT      842 a 1147 c 1136 g 830 t
ORIGIN

alignment_scores:
  Quality: 43.00      Length: 9
  Ratio: 4.778       Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-277-074-10 x RNNEUR

Align seg 1/1 to: RNNEUR from: 1 to: 3955

1 LysillePheGlySerLeuAlaPheLeu 9
|||||
1133 AAGATCTTTGGGACCTGGCATTTTG 1159

seq_name: gb_ro:HAMNEU
seq_documentation_block:
LOCUS      HAMNEU      4062 bp      mRNA      01-FEB-2000

DEFINITION Syrian golden hamster neu mRNA for p-185, complete cds.
ACCESSION D16295
VERSION 1
KEYWORDS D16295.1 GI:493236
Egf-receptor; cellular oncogene; neu-differentiation
factor/hereregulin; p-185; tyrosine kinase.
SOURCE Mesocricetus auratus (individual isolate animal 14) peripheral
nerve neoplastic Schwann cell cell-line 14-2 (library: NIH3T3) cDNA
to mRNA, clone pSHN.
ORGANISM Mesocricetus auratus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
Mesocricetus.
REFERENCE 1 (bases 1 to 4062)
AUTHORS Nakamura,T., Ushijima,T., Ishizaka,Y., Nageo,M., Arai,M.,
Yamazaki,Y. and Ishikawa,T.
TITLE Cloning and activation of the Syrian hamster neu proto-oncogene
JOURNAL Gene 140 (2), 251-255 (1994)
MEDLINE 94193007
REFERENCE 2 (bases 1 to 4062)
AUTHORS Nakamura,T.
TITLE Direct Submission
JOURNAL Submitted (19-MAY-1993) to the DDBJ/EMBL/GenBank databases. Takuro
Nakamura, Faculty of Medicine, University of Tokyo, Department of
Pathology; 7-3-1 Hongo, Bunkyo-ku, Tokyo 113, Japan
(Tel:03-3812-2111(ex.3356), Fax:03-3815-8379)
Submitted (19-May-1993) to DDBJ by:
Takuro Nakamura
Department of Pathology
Faculty of Medicine, University of Tokyo
7-3-1 Hongo
Bunkyo-ku Tokyo 113
Japan
Phone: 03-3812-2111 x3356
Fax: 03-3815-8379.
FEATURES
source
1..4062
/organism="Mesocricetus auratus"
/isolate="animal 14"
/db_xref="taxon:10036"
/cell_line="14-2"
/clone_lib="NIH3T3"
/tissue_type="peripheral nerve"
33..3797
/gene="neu"
33..86
/sig_peptide
/gene="neu"
/product="putative"
33..3797
/gene="neu"
/codon_start=1
/product="p-185 precursor"
/protein_id="BAA03801.1"
/db_xref="GI:747595"
/translation="MELAAWCRWGLALLSPGASGTQVCTGDMKRLRPASPETHLD
IVRHYOCQVVOGNLEIYYPANATLSFLQDIOEVQYMLIAHVSQVRHVPLORLRIV
RGTLQFEDKYALAVLNDPNDVTTATGTPGRLQRLSLTEILKGGVLRGNPO
LCYDPTVLMKDVFKNNQLAPVDIDTNRSRACPPACADKNHCWGESPEDCOITGT
DTFESMNPGRYTFGASCTTCYNYLSTEVGSCITLCPNNOEVFAEDGTQRCEK
SKCARVCYGLGMEHLRGARITSDNOEFGCKIFGSLAFLESFDPGSSGIAPL
TPEOLQVFEETLGYLISAWPDSRLDLSFQNLRIIRGLILHSGAYSILTLQGLGI
RWGLRSRLRGLSGLAIHRNHLCFVHTVPDOLFRNPHQALLHSGNRPESGCLKD
FACYPLCAHGHCHWGPCTQCNCVSHFLRGQECVEECRVKGLPREYNGRKLCPHE
CQPNSTETCTGSEADQCTACPHYKDPFCPCVCPGKVPDLSTFMPLWKYPDEGMCO
PCPINTCHSCVDLDERGPAEQASPVFIATVEGVLFLILVYVVGILIRRRQKI
RKYTMRLQLQETLVEPLTPSGAMPQAQMRILKTELKRVKVGSGAFGTYYKGIWI
PDGENVKIPVAKVLRNTPSKANKELDEAYVMAGLSVPVSRLLGICLTSTVLQVT
OLMXYGCLLDHVRHGRGLSQDLLNWCQIAGKMSYLEDVRLVHRDLAARNVLPKSP
NHVKITDFGLARLLDIDETEHADGCKVPLKWALESILRRBETHQDSVXGYVWE
VSEFSRMARDPQRFVYQNEIDLGSPSLDSTFYRSLLEDDDDMDGLVDABEYLPQOGF
FFPDPAFGAGSTAHRHRRSSSTRSGGELTLMGPEPPSPRSLAPSEGAGSDVFEg
```



```

VERSION      I21124.1  GI:1601478
KEYWORDS
SOURCE       Unknown.
ORGANISM     Unknown.
REFERENCE    1 (bases 1 to 4530)
AUTHORS      Razluddin, J. and Sarkar, F.H.
TITLE        ERBB2 promoter binding protein in neoplastic disease
JOURNAL      Patent: US 5518883-A 9 21-MAY-1996;
FEATURES     Location/Qualifiers
             1..4530
             /organism="unknown"
BASE COUNT   922 a 1382 c 1346 g 880 t
ORIGIN

alignment_scores:
  Quality: 43.00      Length: 9
  Ratio: 4.778       Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
  US-09-277-074-10 x I21124
  Align seg 1/1 to: I21124 from: 1 to: 4530
  1 LysilePheGlySerLeuAlaPheLeu 9
  |||||
  1255 AAGATCTTTGGAGCGCTGGCATTCTG 1281
  seq_name: gb_pat:I59745

seq_documentation_block: 4530 bp DNA PAT
LOCUS      I59745
DEFINITION Sequence 9 from patent US 5654406.
ACCESSION  I59745
VERSION     I59745.1  GI:2478377
KEYWORDS
SOURCE      Unknown.
ORGANISM    Unknown.
REFERENCE    1 (bases 1 to 4530)
AUTHORS      Razluddin, J. and Sarkar, F.Hoque.
TITLE        Antibody to ERBB2 promoter binding factor
JOURNAL      Patent: US 5654406-A 9 05-AUG-1997;
FEATURES     Location/Qualifiers
             1..4530
             /organism="unknown"
BASE COUNT   922 a 1382 c 1346 g 880 t
ORIGIN

alignment_scores:
  Quality: 43.00      Length: 9
  Ratio: 4.778       Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
  US-09-277-074-10 x I59745
  Align seg 1/1 to: I59745 from: 1 to: 4530
  1 LysilePheGlySerLeuAlaPheLeu 9
  |||||
  1255 AAGATCTTTGGAGCGCTGGCATTCTG 1281
  seq_name: gb_pr7:HUMHER2A

seq_documentation_block:
LOCUS      HUMHER2A 4530 bp mRNA PRI
DEFINITION Human tyrosine kinase-type receptor (HER2), complete cds.
ACCESSION  M11730
VERSION     M11730.1  GI:183986

```

---

```

KEYWORDS     tyrosine kinase.
SOURCE       Homo sapiens (Clone: lambda-HER2-436) fetal cDNA to mRNA.
ORGANISM     Homo sapiens
REFERENCE    1 (bases 1 to 4530); 1 to 4530
AUTHORS      Cousens, L., Yang-Feng, T.L., Liao, Y.-C., Chen, E., Gray, A.,
              McGrath, J., Seeburg, P.H., Libermann, T.A., Schlessinger, J.,
              Francke, U., Levinson, A. and Ullrich, A.
TITLE        Tyrosine kinase receptor with extensive homology to EGF receptor
              shares chromosomal location with neu oncogene
JOURNAL      Science 230 (4730), 1132-1139 (1985)
FEATURES     Location/Qualifiers
             1..4530
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             /clone="lambda-HER2-436"
             /dev_stage="fetal"
             <1..4530
             /note="HER2 mRNA"
             151..3918
             /note="HER2 receptor"
             /codon_start=1
             /protein_id="AA75493.1"
             /translation="MELAALCRWGLLLALLPPGAASQTVCCTGMKRLPASPETHLD
              MRLHYQGVVQVQNLTYLTPTNASLSFLQIQEVQVYVLIHNRVQVPLQRRIY
              RGTLQFEDNYALAVLDNGPLNPTPTVGASPGRLQRLSLTEILKGGVLIQRNPQ
              LCVQDTILKDIHFKNOLALTLDITNRSRACHPCSPCKSGRSGESSEDCSLTTRT
              VCAGCARCKGFLPTDCHEGCAAGCTGPKHSDCLACLHFNHSGICEHCPALVTNT
              DTFESMPNPEGRITGASCVTACPNISTDVGSCITLVCLPLNHQVEVDGDPASNTAPL
              SKPCARQVETGIEITGYLIISAMPDLSPLSVFQNLQYIRGLHNGAYSLTQGLGI
              QPEQLQVFTLEITGYLIISAMPDLSPLSVFQNLQYIRGLHNGAYSLTQGLGI
              SWGLRSLRELGSGLAIHNTHLCFVHTVMDQLFRNPQHALLHNRNPDECEVGE
              LACHOLCARGHCWGPGTCVNCQFLRGQECVEFCVQLQGLPREYVNRHCLPCGHE
              COPNGSVTCFGEADQCVACAHYKDPFCVACRCPGKVDLSYVPIKPFDEGACQ
              PCPINCTHSCVDLDDKGCFAQRASPLTSIVSAVVGILLVVGIVGVFGLLIRKROKI
              RKYTMRLQLQETELVEPLTPSGAMPNQMRILKTELKRVKVLGSGAFGTVYKGIWI
              PDGENVKIPVAIKVLRNTSPKANKEILDEAYVMAGVSPYVSRLLGCLTSTVQLVT
              OLMPYKCLLDHVRNRLGSLQDLNMCQIAKMSYLEDVRLHRLGLAARNLVKSP
              NHVKITDFGLARLLDIDTEYHADGKVKPIKMALESLIRRFETHQSDWSYGVTVWE
              LMTFGAKPYDGIIPAREIDPLEKGERLPQPPCTIDYVIMVKNMIDSECRPREL
              VSEFSRMARDQRFVYVIONEDLGPASPLDFTYRSLDDDDMDGLVDAAEYLYPQGGF
              FCPDPAPGAGGVHHRSSSTRSGGDLTLGLEPSEEAAPRLAPSEAGSDVFDG
              DLGMAAKGLASLPTHPDPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYVNPQDVR
              POPPSREGPLPAARPAAGATLERAKTLSPKNGVYKDFAFGGAVENPEYLTPOGGAA
              PQHPFPAPFAPFDNLYWDQDPPPERGAPPSTFKCTPTAENPEYLGLDVVPV"
             1701..1719
             /citation=[1]
BASE COUNT   922 a 1382 c 1346 g 880 t
ORIGIN       Chromosome 17q21-q22.

alignment_scores:
  Quality: 43.00      Length: 9
  Ratio: 4.778       Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
  US-09-277-074-10 x HUMHER2A
  Align seg 1/1 to: HUMHER2A from: 1 to: 4530
  1 LysilePheGlySerLeuAlaPheLeu 9
  |||||
  1255 AAGATCTTTGGAGCGCTGGCATTCTG 1281
  seq_name: gb_htg11:AC025531

```

```

seq_documentation_block:
LOCUS      AC025531 154278 bp  DNA           HTG           07-JUL-2000
DEFINITION Homo sapiens chromosome 17 clone RP11-62N23, WORKING DRAFT
SEQUENCE, 24 unchromed pieces.
ACCESSION  AC025531
VERSION    AC025531.3 GI:8569754
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 154278)
            Waterston,R.H.
            The sequence of Homo sapiens clone
            Unpublished
REFERENCE  2 (bases 1 to 154278)
            Waterston,R.H.
            Direct Submission
            Submitted (10-MAR-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT    On Jun 16, 2000 this sequence version replaced gi:7230850.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_MH0062N23
----- Summary Statistics -----
Sequencing vector: M13; 100%
Sequencing vector: plasmid; 0%
Chemistry: Dye-primer ET; 100% of reads
Chemistry: Dye-terminator Big Dye; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 136833 bases at least Q40
Consensus quality: 143642 bases at least Q30
Consensus quality: 147431 bases at least Q20
Insert size: 155000; agarose-fp
Insert size: 151978; sum-of-contigs
Quality coverage: 3.51 in Q20 bases; agarose-fp
Quality coverage: 3.62 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 24 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1107: contig of 1107 bp in length
* 1108 1207: gap of unknown length
* 1208 2357: contig of 1150 bp in length
* 2358 2457: gap of unknown length
* 2458 4301: contig of 1844 bp in length
* 4302 4401: gap of unknown length
* 4402 6082: contig of 1681 bp in length
* 6083 6183: gap of unknown length
* 6183 7579: contig of 1397 bp in length
* 7580 7679: gap of unknown length
* 7680 10360: contig of 2681 bp in length
* 10361 10460: gap of unknown length
* 10461 12061: contig of 1601 bp in length
* 12062 12161: gap of unknown length
* 12162 15340: contig of 3179 bp in length
* 15341 15441: gap of unknown length
* 15441 17469: contig of 2029 bp in length
* 17470 17570: gap of unknown length
* 17570 20933: contig of 3364 bp in length
* 20934 21034: gap of unknown length
* 21034 23604: contig of 2571 bp in length
* 23605 23704: gap of unknown length

```

```

* 23705 27949: contig of 4245 bp in length
* 27950 28049: gap of unknown length
* 28050 33165: contig of 5116 bp in length
* 33166 33265: gap of unknown length
* 33266 39025: contig of 5760 bp in length
* 39026 39125: gap of unknown length
* 39126 46324: contig of 7199 bp in length
* 46325 46424: gap of unknown length
* 46425 54131: contig of 7707 bp in length
* 54132 54231: gap of unknown length
* 54232 59390: contig of 5159 bp in length
* 59391 59490: gap of unknown length
* 59491 69238: contig of 9748 bp in length
* 69239 69338: gap of unknown length
* 69339 78047: contig of 8709 bp in length
* 78048 78147: gap of unknown length
* 78148 89237: contig of 11090 bp in length
* 89238 89337: gap of unknown length
* 89338 103171: contig of 13834 bp in length
* 103172 103271: gap of unknown length
* 103272 119511: contig of 16240 bp in length
* 119512 119611: gap of unknown length
* 119612 133144: contig of 13533 bp in length
* 133145 133244: gap of unknown length
* 133245 154278: contig of 21034 bp in length.

FEATURES
            Location/Qualifiers
            1..154278
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="17"
            /clone="RP11-62N23"
            1..1107
            /note="assembly_name:Contig7"
            1208..2357
            /note="assembly_name:Contig12"
            2458..4301
            /note="assembly_name:Contig13"
            4402..6082
            /note="assembly_name:Contig14"
            6183..7579
            /note="assembly_name:Contig15"
            7680..10360
            /note="assembly_name:Contig16"
            10461..12061
            /note="assembly_name:Contig17"
            12162..15340
            /note="assembly_name:Contig18"
            15441..17469
            /note="assembly_name:Contig19"
            17570..20933
            /note="assembly_name:Contig20"
            21034..23604
            /note="assembly_name:Contig21"
            clone_end:r7
            vector_side:right
            23705..27949
            /note="assembly_name:Contig22"
            28050..33165
            /note="assembly_name:Contig23"
            33266..39025
            /note="assembly_name:Contig24"
            39126..46324
            /note="assembly_name:Contig25"
            46425..54131
            /note="assembly_name:Contig26"
            54232..59390
            /note="assembly_name:Contig27"
            clone_end:sp6
            vector_side:left
            59491..69238
            /note="assembly_name:Contig28"
            69339..78047
            /note="assembly_name:Contig29"

```

misc\_feature 78148..89237  
 /note="assembly\_name:Contig30"  
 misc\_feature 89338..103171  
 /note="assembly\_name:Contig31"  
 misc\_feature 103272..119511  
 /note="assembly\_name:Contig32"  
 misc\_feature 119612..133144  
 /note="assembly\_name:Contig33"  
 misc\_feature 133245..154278  
 /note="assembly\_name:Contig34"  
 BASE COUNT 34471 a 40782 c 41099 g 35604 t 2322 others  
 ORIGIN

alignment\_scores:  
 Quality: 43.00 Length: 9  
 Ratio: 4.778 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
 US-09-277-074-10 x AC025531/rev ..  
 Align seg 1/1 to reverse of: AC025531 from: 1 to: 154278

1 LysilePheGlySerLeuAlaPheLeu 9  
 |||||  
 68216 AGATCTTTGGACCGTCGATTCCTG 68190

seq\_name: gb\_v12:RRNASEG3

seq\_documentation\_block:  
 LOCUS RRNASEG3 1318 bp RNA VRL 17-OCT-1995  
 DEFINITION Reovirus sp. 1.318kb RNA segment.  
 ACCESSION X82047  
 VERSION X82047.1 GI:1022337  
 KEYWORDS  
 SOURCE Reovirus sp.  
 ORGANISM Reovirus sp.  
 REFERENCE 1 (bases 1 to 1318)  
 AUTHORS Bigot,Y.  
 TITLE Direct Submission  
 JOURNAL Submitted (04-OCT-1994) Y. Bigot, IBEAS, Faculte des Sciences, Parc Grandmont, Ave. Monge, 37200 Tours, FRANCE  
 REFERENCE 2 (bases 1 to 1318)  
 AUTHORS Bigot,Y., Drezen,J.M., Sizaret,P.Y., Rabouille,A., Hamelin,M.H. and Periquet,G.  
 TITLE The genome segments of DpRV, a commensal reovirus of the wasp  
 JOURNAL Diadromus pulchellus (Hymenoptera)  
 MEDLINE Virology 210 (1), 109-119 (1995)  
 FEATURES  
 Location/Qualifiers  
 1..1318  
 /organism="Reovirus sp."  
 /specific\_host="Diadromus pulchellus"  
 /db\_xref="taxon:10891"  
 53..1252  
 /gene="ORF"  
 53..1252  
 /gene="ORF"  
 /codon\_start=1  
 /protein\_id="CAA57563.1"  
 /db\_xref="GI:1022338"  
 /db\_xref="SPTREMBL:Q86284"  
 /translation="NNPEKARQKLKKNAKRNLTILNMESTPEAQSVDNISQSQ  
 NDLTKLTVPKPSISTKILKRKLTISDLIKLFTCDVHEDDIYVGGSKSDFT  
 IKKLYKANVPKLYTNVYVHTDDHIGELASERSKRLMOTLSLVDINKYDQES  
 QMTLSMKQACIEVKKQIPLAQENDTNVERILSGNSASSVSECEQDYMDEQS  
 ADGNERVSEIIDLNAVVPVLNDMDSTESGDIGTELISMFAGQMEFFSPPLAKE  
 NMPEDVLEEMKLLDELEARYKMTASTNKVTKLLQSAHNYQGFSDNKRNLISR  
 INDLNPLIEYADKIDNFCOLYGPADQPIAIGRVINIAKRVISVNGLTNKLKNG  
 KFLTNAE"

BASE COUNT 519 a 243 c 259 g 297 t

# ORIGIN

alignment\_scores:  
 Quality: 40.00 Length: 9  
 Ratio: 4.444 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 88.889  
 alignment\_block:  
 US-09-277-074-10 x RRNASEG3/rev ..  
 Align seg 1/1 to reverse of: RRNASEG3 from: 1 to: 1318

1 LysilePheGlySerLeuAlaPheLeu 9  
 |||||  
 402 AAATTTTCGTACATGGCATCTTG 376

seq\_name: gb\_pri:AC004998

seq\_documentation\_block:  
 LOCUS AC004998 135572 bp DNA PRI 21-DEC-1999  
 DEFINITION Homo sapiens PAC clone RPL-164D5 from Xq23, complete sequence.  
 ACCESSION AC004998  
 VERSION AC004998.2 GI:5091651  
 KEYWORDS HTG.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 135572)  
 AUTHORS Sulston,J.E. and Waterston,R.  
 TITLE Toward a complete human genome sequence  
 JOURNAL Genome Res. 8 (11), 1097-1108 (1998)  
 MEDLINE 99063792  
 REFERENCE 2 (bases 1 to 135572)  
 AUTHORS Andrews,S.  
 TITLE The sequence of Homo sapiens PAC clone RPL-164D5  
 JOURNAL Unpublished  
 REFERENCE 3 (bases 1 to 135572)  
 AUTHORS Waterston,R.H.  
 TITLE Direct Submission  
 JOURNAL Submitted (12-JUN-1998) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

REFERENCE 4 (bases 1 to 135572)  
 AUTHORS Waterston,R.H.  
 TITLE Direct Submission  
 JOURNAL Submitted (17-JUN-1999) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

REFERENCE 5 (bases 1 to 135572)  
 AUTHORS Waterston,R.  
 TITLE Direct Submission  
 JOURNAL Submitted (21-DEC-1999) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
 On Jun 17, 1999 this sequence version replaced gi:3212993.

----- Genome Center  
 Center: Washington University Genome Sequencing Center  
 Center code: WUGSC  
 Web site: http://genome.wustl.edu/gsc  
 Contact: sapiens@watson.wustl.edu  
 ----- Summary Statistics  
 -----  
 Center project name: H\_DJ164D05

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
 all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >=

30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

#### MAPPING INFORMATION:

This sequence was generated from part of bacterial clone contigs of human chromosome X, constructed by the chromosome X mapping group at the Sanger Centre, Wellcome Trust Genome Campus, Hinxton, UK. Further information can be found at <http://www.sanger.ac.uk/HGP/ChrX/>

#### SOURCE INFORMATION:

This clone was derived from human PAC library RPI-1, prepared by Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://pacpac.med.buffalo.edu>) using the method described by Ioannou et al., Nature Genetics 6:84-9 (1994). The library is from one male donor. The clone may be obtained either from Genome Systems, Inc. (<http://www.genomesystems.com>) or Research Genetics, Inc. (<http://www.resgen.com>); or from Pieter de Jong. VECTOR: pCYPAC2

#### NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RPI-73F11, 200 bp overlap. Actual start of this clone is at base position 83380 of RPI-73F11; actual end is at base position 135572 of RPI-164D5.

#### FEATURES

source	Location/Qualifiers
repeat_region	1. 135572
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="X"
	/map="Xq23"
	/clone="RPI-164D5"
	/clone_lib="RPI-1"
	1388. 1427
repeat_region	/rpt_family="(CA)n"
repeat_region	1540. 1674
	/rpt_family="MER21-g"
repeat_region	1885. 1895
	/rpt_family="Alu"
repeat_region	1896. 1919
	/rpt_family="(CAAAA)n"
repeat_region	1920. 2200
	/rpt_family="Alu"
repeat_region	2225. 2604
	/rpt_family="MaLR"
repeat_region	2684. 2757
	/rpt_family="MER21-g"
repeat_region	3333. 3543
	/rpt_family="MIR"
repeat_region	3614. 3723
	/rpt_family="L2"
repeat_region	3728. 4015
	/rpt_family="MER4-group"
repeat_region	4008. 4339
	/rpt_family="MER4-group"
repeat_region	4533. 4656
	/rpt_family="L2"
repeat_region	5735. 5945
	/rpt_family="MIR"
repeat_region	6127. 6439
	/rpt_family="Alu"
repeat_region	6543. 7037
	/rpt_family="MER21-g"
repeat_region	7845. 7879
	/rpt_family="(CAT)n"
repeat_region	11373. 11466
	/rpt_family="L2"
repeat_region	11561. 11784
	/rpt_family="MIR"
repeat_region	12187. 12296
	/rpt_family="L1"
repeat_region	12433. 12465
	/rpt_family="AT_rich"
repeat_region	12907. 12949
	/rpt_family="(GAAA)n"
repeat_region	13527. 13627
	/rpt_family="L2"
repeat_region	13667. 14237
	/rpt_family="L1"
repeat_region	14238. 14540
	/rpt_family="Alu"
repeat_region	14541. 14936
	/rpt_family="L1"
repeat_region	14937. 15086
	/rpt_family="Alu"
repeat_region	15087. 15614
	/rpt_family="L1"
repeat_region	15615. 15914
	/rpt_family="Alu"
repeat_region	15915. 16546
	/rpt_family="L1"
repeat_region	16514. 16751
	/rpt_family="L1"
repeat_region	16845. 19911
	/rpt_family="L1"
repeat_region	19912. 20180
	/rpt_family="Alu"
repeat_region	20181. 24034
	/rpt_family="L1"
repeat_region	24063. 24100
	/rpt_family="Alu"
repeat_region	24102. 24404
	/rpt_family="Alu"
repeat_region	24406. 24583
	/rpt_family="L1"
repeat_region	24584. 24897
	/rpt_family="L1"
repeat_region	24971. 25135
	/rpt_family="L1"
repeat_region	25203. 25312
	/rpt_family="L2"
repeat_region	25880. 26153
	/rpt_family="Alu"
repeat_region	26190. 26281
	/rpt_family="AT_rich"
repeat_region	26365. 26386
	/rpt_family="(CAAAA)n"
repeat_region	26659. 26947
	/rpt_family="Alu"
repeat_region	26948. 27004
	/rpt_family="AT_rich"
repeat_region	27219. 27252
	/rpt_family="AT_rich"
repeat_region	27345. 28388
	/rpt_family="L1"
repeat_region	28389. 28418
	/rpt_family="(TAAA)n"
repeat_region	28952. 28986
	/rpt_family="AT_rich"
repeat_region	29240. 30872
	/rpt_family="L1"
repeat_region	30974. 31956
	/rpt_family="L1"
repeat_region	32696. 32785
	/rpt_family="MaLR"
repeat_region	33142. 33167
	/rpt_family="(TAA)n"
repeat_region	33417. 33694
	/rpt_family="Alu"
repeat_region	33695. 33727
	/rpt_family="(TAA)n"
repeat_region	34115. 34487
	/rpt_family="L2"
repeat_region	34517. 34544
	/rpt_family="AT_rich"



Thu Nov 16 15:46:32 2000

```

repeat_region 34589..34719

alignment_scores:
  Quality: 40.00      Length: 9
  Ratio: 4.44         Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 88.889

alignment_block:
  US-09-277-074-10 x AC004998/rev ..

  Align seg 1/1 to reverse of: AC004998 from: 1 to: 135572

    1 LyslePheGlySerLeuAlaphelou 9
      |||||
  59165 AAATATTGGCTCATTAATCTCTTG 59139

seq_name: gb_htg21:AL359545

seq_documentation_block:
  LOCUS AL359545 160731 bp DNA HTG 15-JUL-2000
  DEFINITION Homo-sapiens chromosome X clone RP13-42E14, *** SEQUENCING IN
  PROGRESS ***, 37 unordered pieces.
  ACCESSION AL359545
  VERSION AL359545.5 GI:9367612
  KEYWORDS HTG: HTGS_PHASE1.
  SOURCE human.
  ORGANISM Homo sapiens
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  1 (bases 1 to 160731)
  Pavitt R.
  Direct Submission
  Submitted (15-JUL-2000) Sanger Centre, Hinxton, Cambridgeshire,
  CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
  requests: clonerequest@sanger.ac.uk
  On Jul 22, 2000 this sequence version replaced gi:9187272.
  ----- Genome Center
  Center: Sanger Centre
  Center code: SC
  Web site: http://www.sanger.ac.uk
  Contact: humquery@sanger.ac.uk
  ----- Project Information
  Center project name: bB42E14
  ----- Summary Statistics
  Assembly program: XGAP4; version 4.5
  Sequencing vector: plasmid; L08752; 100% of reads
  Chemistry: Dye-terminator ET-amersham; 100% of reads
  Consensus quality: 147707 bases at least Q40
  Consensus quality: 151841 bases at least Q30
  Consensus quality: 154336 bases at least Q20
  Insert size: 157131; sum-of-contigs
  Insert size: 171884; 3.0% error; agarose-fp
  Quality coverage: 3.38x in Q20 bases; sum-of-contigs Quality
  coverage: 3.09x in Q20 bases; agarose-fp
  -----
  * NOTE: This is a 'working draft' sequence. It currently
  * consists of 37 contigs. The true order of the pieces
  * is not known and their order in this sequence record is
  * arbitrary. Gaps between the contigs are represented as
  * runs of N, but the exact sizes of the gaps are unknown.
  * This record will be updated with the finished sequence.
  * as soon as it is available and the accession number will
  * be preserved.
  *
  * 1 3111: contig of 3111 bp in length
  * 3112 3211: gap of 100 bp
  * 3212 4238: contig of 1017 bp in length
  * 4239 4328: gap of 100 bp
  * 4329 14914: contig of 10586 bp in length
  * 14915 15014: gap of 100 bp
  * 15015 16326: contig of 1312 bp in length
  * 16327 16426: gap of 100 bp
  * 16427 24672: contig of 8245 bp in length
  * 24673 24772: gap of 100 bp
  *
  FEATURES
    source
      1..160731
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /chromosome="X"
        /clone="RP13-42E14"
        /clone_lib="RPCI-13.1"
        1..3111
        /note="assembly_fragment:01186
        fragment_chain:1
    misc_feature
      1..3111
        /note="assembly_fragment:01186
        fragment_chain:1

```

misc\_feature clone\_end:SP6  
vector\_side:left  
3212..4228  
/note="assembly\_fragment:00865  
fragment\_chain:1"  
4329..14914  
/note="assembly\_fragment:01165  
fragment\_chain:1"  
15015..16326  
/note="assembly\_fragment:00302  
fragment\_chain:1"  
16427..24672  
/note="assembly\_fragment:00784  
fragment\_chain:1"  
24773..29912  
/note="assembly\_fragment:00488  
fragment\_chain:1"  
30013..34181  
/note="assembly\_fragment:00219  
fragment\_chain:1"  
34282..38024  
/note="assembly\_fragment:01146  
fragment\_chain:1"  
38125..39143  
/note="assembly\_fragment:00572  
fragment\_chain:2"  
39244..41371  
/note="assembly\_fragment:00426  
fragment\_chain:2"  
41472..47697  
/note="assembly\_fragment:00630  
fragment\_chain:2"  
47798..50653  
/note="assembly\_fragment:00668  
fragment\_chain:3"  
50754..52744  
/note="assembly\_fragment:00983  
fragment\_chain:3"  
62845..69524  
/note="assembly\_fragment:00477  
fragment\_chain:3"  
69625..73840  
/note="assembly\_fragment:00154  
fragment\_chain:4"  
73941..75434  
/note="assembly\_fragment:00454  
fragment\_chain:4"  
75535..80493  
/note="assembly\_fragment:00533  
fragment\_chain:5"  
80594..82445  
/note="assembly\_fragment:00514  
fragment\_chain:5"  
82546..85046  
/note="assembly\_fragment:00694  
fragment\_chain:6"  
85147..89613  
/note="assembly\_fragment:00597  
fragment\_chain:6"  
89714..91617  
/note="assembly\_fragment:01116  
fragment\_chain:7"  
91718..93751  
/note="assembly\_fragment:00044  
fragment\_chain:7"  
95852..101065  
/note="assembly\_fragment:00088"  
101166..109849  
/note="assembly\_fragment:00152"  
109950..114489  
/note="assembly\_fragment:00164"  
114590..121031  
/note="assembly\_fragment:00225"

alignment\_scores:  
Quality: 39.00 Length: 9  
Ratio: 4.333 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 88.889  
alignment\_block:  
US-09-277-074-10 x AL359545 ..  
Align seg 1/1 to: AL359545 from: 1 to: 160731  
1 LysilepHeGlySerLeuAlapheLeu 9  
|||||  
154143 AATAATTTGGAGCTGGTTTCTTG 154169  
seq\_name: gb\_htg21:AL359548  
seq\_documentation\_block:  
LOCUS AL359548 166012 bp DNA HTG 20-JUN-2000  
DEFINITION Homo sapiens chromosome X clone RP11-167F1, \*\*\* SEQUENCING IN  
PROGRESS \*\*\*, 26 unordered pieces.  
ACCESSION AL359548  
VERSION AL359548.2 GI:9231227  
KEYWORDS HTG: HTGS\_PHASE1.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
AUTHORS Pavitt,R.  
TITLE Direct Submission  
JOURNAL Submitted (17-JUN-2000) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
requests: clonerequest@sanger.ac.uk  
On Jul 16, 2000 this sequence version replaced gi:8670695.  
COMMENT  
----- Genome Center  
Center: Sanger Centre  
Center code: SC  
Web site: <http://www.sanger.ac.uk>  
Contact: humquery@sanger.ac.uk  
----- Project Information  
Center project name: BA167F1  
----- Summary Statistics  
Assembly program: XGAP4; version 4.5  
Sequencing vector: plasmid; L08752; 100% of reads  
Chemistry: Dye-terminator Big Dye; 38% of reads  
Chemistry: Dye-terminator ET-amersham; 61% of reads  
Consensus quality: 154071 bases at least Q40  
Consensus quality: 158958 bases at least Q30  
Consensus quality: 161761 bases at least Q20  
Insert size: 163512; sum-of-contigs  
Insert size: 125203; 41.8% error; agarose-fp  
Quality coverage: 3.54x in Q20 bases; sum-of-contigs Quality  
coverage: 4.63x in Q20 bases; agarose-fp  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 26 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 11670: contig of 11670 bp in length  
\* 11671 11770: gap of 100 bp  
\* 11771 21158: contig of 9388 bp in length  
\* 21159 21258: gap of 100 bp  
\* 21259 23527: contig of 2269 bp in length  
\* 23528 23627: gap of 100 bp  
\* 23628 27914: contig of 4287 bp in length  
\* 27915 28014: gap of 100 bp  
\* 28015 30783: contig of 2769 bp in length  
\* 30784 30883: gap of 100 bp

```
* 30884 41169: contig of 10286 bp in length
* 41170 41269: gap of 100 bp
* 41270 42838: contig of 1569 bp in length
* 42839 42938: gap of 100 bp
* 42939 49762: contig of 6824 bp in length
* 49763 49862: gap of 100 bp
* 49863 51422: contig of 1560 bp in length
* 51423 51522: gap of 100 bp
* 51523 78695: contig of 27173 bp in length
* 78696 78795: gap of 100 bp
* 78796 83185: contig of 4390 bp in length
* 83186 83285: gap of 100 bp
* 83286 84429: contig of 1144 bp in length
* 84430 84529: gap of 100 bp
* 84530 90854: contig of 6325 bp in length
* 90855 90954: gap of 100 bp
* 90955 97762: contig of 6808 bp in length
* 97763 97862: gap of 100 bp
* 97863 100197: contig of 2335 bp in length
* 100198 100297: gap of 100 bp
* 100298 105155: contig of 4858 bp in length
* 105156 105253: gap of 100 bp
* 105254 118475: contig of 13220 bp in length
* 118476 118575: gap of 100 bp
* 118576 122442: contig of 3867 bp in length
* 122443 123543: gap of 100 bp
* 123544 124038: contig of 1496 bp in length
* 124039 124139: gap of 100 bp
* 124140 132453: contig of 8315 bp in length
* 132454 132553: gap of 100 bp
* 132554 135187: contig of 2634 bp in length
* 135188 135287: gap of 100 bp
* 135288 146152: contig of 10865 bp in length
* 146153 146252: gap of 100 bp
* 146253 147953: contig of 1701 bp in length
* 147954 148053: gap of 100 bp
* 148054 162005: contig of 13952 bp in length
* 162006 162105: gap of 100 bp
* 162106 163248: contig of 1143 bp in length
* 163249 163349: gap of 100 bp
* 163349 166012: contig of 2664 bp in length.
FEATURES
    source
    1..166012
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone="RP11-167F1"
    /chromosome="X"
    /clone_lib="RPC1-11.1"
    1..11670
    /note="assembly_fragment:01085"
    clone_end:SP6
    vector_side:left
    11771..21158
    /note="assembly_fragment:01024"
    fragment_chain:1"
    21259..23527
    /note="assembly_fragment:00073"
    fragment_chain:1"
    23628..27914
    /note="assembly_fragment:00027"
    fragment_chain:1"
    28015..30783
    /note="assembly_fragment:01042"
    fragment_chain:1"
    30884..41169
    /note="assembly_fragment:00928"
    fragment_chain:1"
    41270..42838
    /note="assembly_fragment:00577"
    fragment_chain:2"
    42939..49762
    /note="assembly_fragment:00203"
    fragment_chain:2"
```

```
misc_feature 49863..51422
              /note="assembly_fragment:00329"
              fragment_chain:2"
misc_feature 51523..78695
              /note="assembly_fragment:00131"
              fragment_chain:2"
misc_feature 78796..83185
              /note="assembly_fragment:00075"
              fragment_chain:3"
misc_feature 83286..84429
              /note="assembly_fragment:01272"
              fragment_chain:3"
misc_feature 84530..90854
              /note="assembly_fragment:00722"
              fragment_chain:4"
misc_feature 90955..97762
              /note="assembly_fragment:01236"
              fragment_chain:4"
misc_feature 97863..100197
              /note="assembly_fragment:00249"
              fragment_chain:4"
misc_feature 100298..105155
              /note="assembly_fragment:00654"
              fragment_chain:4"
misc_feature 105256..118475
              /note="assembly_fragment:00684"
              fragment_chain:4"
misc_feature 118576..122442
              /note="assembly_fragment:00878"
              fragment_chain:4"
misc_feature 122543..124038
              /note="assembly_fragment:00920"
              fragment_chain:4"
misc_feature 124139..132453
              /note="assembly_fragment:00968"
              fragment_chain:4"
misc_feature 132554..135187
              /note="assembly_fragment:00983"
              fragment_chain:4"
misc_feature 135288..146152
              /note="assembly_fragment:01026"
              fragment_chain:4"
misc_feature 146253..147953
              /note="assembly_fragment:01245"
              fragment_chain:4"
misc_feature 148054..162005
              /note="assembly_fragment:01387"
              fragment_chain:4"
misc_feature 162106..163248
              /note="assembly_fragment:01464"
              fragment_chain:4"
misc_feature 163349..166012
              /note="assembly_fragment:00857"
              clone_end:T7
              vector_side:right"
BASE COUNT 49752 a 30517 c 30813 g 52424 t 2506 others
ORIGIN
```

```
alignment_scores:
    Quality: 39.00          Length: 9
    Ratio: 4.333           Gaps: 0
Percent Similarity: 100.000 Percent Identity: 88.889

alignment_block:
US-09-277-074-10 x AL359548 ..
Align seg 1/1 to: AL359548 from: 1 to: 166012
1 LysilePheGlySerLeuAlaPheLeu 9
|||||
154327 AAAATATTGGAAGCTGGTTTCTTG 154353
```



OM of: US-09-277-074-10 to: N\_Geneseq\_36.\* out\_format : pfs  
 Date: Nov 15, 2000 10:29 PM  
 About: Results were produced by the GenCore software, version 4.5,  
 Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:  
 -MODEL-frame+2n.model -DEV-xlp  
 -Q/cyn2\_1/USPro\_spool/US09277074/runat\_14112000\_120309\_22640/app\_query.fasta\_1.67  
 -DB-N\_Geneseq\_36 -OFMT-fastap -SUFFIX-rng -GAPOP=12.000  
 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000  
 -GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500  
 -GAPOP=6.000 -GAPEXT=7.000 -XGAPOP=10.000 -XGAPEXT=0.500  
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX-blosum62  
 -TRANS-human40.cdi -LIST=45 -DOCALLIGN=200 -THR\_SCORE=pet  
 -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfis  
 -NORM-ext -MINLEN=0 -MAXLEN=200000000  
 -USER=US09277074 -CGN1\_1\_108 -NCPU=6 -ICPU=3 -LONGLOG -NO\_XLPXY  
 -WAIT -THREADS=1

Search information block:  
 Query: US-09-277-074-10  
 Query length: 9  
 Database: N\_Geneseq\_36.\*  
 Database sequences: 480022  
 Database length: 187831343  
 Search time (sec): 82.090000

score\_list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
/SID56/gcgdata/geneseq/geneseq/NA1990.DAT:Q06828 +		43.00	141.52	1872	I
/SID56/gcgdata/geneseq/geneseq/NA1997.DAT:T72725 +		43.00	139.42	2385	I
/SID56/gcgdata/geneseq/geneseq/NA2000.DAT:Z50586 +		43.00	137.81	2871	I
/SID56/gcgdata/geneseq/geneseq/NA1996.DAT:T40739 +		43.00	135.45	3768	I
/SID56/gcgdata/geneseq/geneseq/NA1999.DAT:X01912 +		43.00	135.45	3768	I
/SID56/gcgdata/geneseq/geneseq/NA2000.DAT:A09455 +		43.00	135.45	3768	I
/SID56/gcgdata/geneseq/geneseq/NA1995.DAT:T01590 +		43.00	135.03	3955	I
/SID56/gcgdata/geneseq/geneseq/NA2000.DAT:A14812 +		43.00	133.97	4472	I
/SID56/gcgdata/geneseq/geneseq/NA1999.DAT:Z31071 +		43.00	133.86	4473	I
/SID56/gcgdata/geneseq/geneseq/NA1997.DAT:T01585 +		43.00	133.86	4530	I
/SID56/gcgdata/geneseq/geneseq/NA1997.DAT:T71253 +		43.00	133.86	4530	I
/SID56/gcgdata/geneseq/geneseq/NA2000.DAT:Z60815 +		43.00	133.86	4530	I
/SID56/gcgdata/geneseq/geneseq/NA1998.DAT:Q46083 +		40.00	123.01	4299	I
/SID56/gcgdata/geneseq/geneseq/NA1993.DAT:V62286 +		37.00	124.62	969	I
/SID56/gcgdata/geneseq/geneseq/NA1998.DAT:V52288 +		35.00	108.15	6474	I
/SID56/gcgdata/geneseq/geneseq/NA1998.DAT:V40501 +		34.00	115.90	719	I
/SID56/gcgdata/geneseq/geneseq/NA1992.DAT:Q20684 +		34.00	106.40	2151	I
/SID56/gcgdata/geneseq/geneseq/NA1992.DAT:Q20759 +		34.00	105.71	2330	I
/SID56/gcgdata/geneseq/geneseq/NA1999.DAT:X13726 +		33.00	109.44	981	I
/SID56/gcgdata/geneseq/geneseq/NA1997.DAT:T98726 +		33.00	107.75	1191	I
/SID56/gcgdata/geneseq/geneseq/NA1997.DAT:T96835 +		33.00	105.10	1618	I
/SID56/gcgdata/geneseq/geneseq/NA1998.DAT:Z92422 +		33.00	105.01	1634	I
/SID56/gcgdata/geneseq/geneseq/NA1998.DAT:V37360 +		33.00	105.01	1634	I
/SID56/gcgdata/geneseq/geneseq/NA1998.DAT:V62221 +		33.00	101.39	2483	I
/SID56/gcgdata/geneseq/geneseq/NA2000.DAT:A08514 +		33.00	99.58	3057	I
/SID56/gcgdata/geneseq/geneseq/NA1998.DAT:V52358 +		33.00	97.77	3766	I
/SID56/gcgdata/geneseq/geneseq/NA1999.DAT:V08862 +		32.00	113.10	416	I
/SID56/gcgdata/geneseq/geneseq/NA1999.DAT:V89881 +		32.00	112.82	430	I
/SID56/gcgdata/geneseq/geneseq/NA1999.DAT:V87411 +		32.00	112.20	462	I
/SID56/gcgdata/geneseq/geneseq/NA1997.DAT:T67605 +		32.00	109.80	609	I
/SID56/gcgdata/geneseq/geneseq/NA2000.DAT:Z79987 +		32.00	108.53	705	I
/SID56/gcgdata/geneseq/geneseq/NA1997.DAT:Z16617 +		32.00	107.77	770	I
/SID56/gcgdata/geneseq/geneseq/NA1997.DAT:T67837 +		32.00	107.39	804	I
/SID56/gcgdata/geneseq/geneseq/NA1986.DAT:N60556 +		32.00	103.06	1325	I
/SID56/gcgdata/geneseq/geneseq/NA1986.DAT:N60555 +		32.00	102.73	1376	I
/SID56/gcgdata/geneseq/geneseq/NA1990.DAT:Q05809 +		32.00	102.73	1376	I
/SID56/gcgdata/geneseq/geneseq/NA1992.DAT:Q23216 +		32.00	102.73	1376	I
/SID56/gcgdata/geneseq/geneseq/NA1999.DAT:X57357 +		32.00	102.59	1399	I

/SID56/gcgdata/geneseq/geneseq/NA1989.DAT:N90120 +	32.00	99.79	905.20	1933
/SID56/gcgdata/geneseq/geneseq/NA1998.DAT:X14080 +	32.00	98.51	1.1e+03	2241
/SID56/gcgdata/geneseq/geneseq/NA1999.DAT:X13400 +	32.00	98.01	1.1e+03	2373
/SID56/gcgdata/geneseq/geneseq/NA1999.DAT:X178785 +	32.00	96.91	1.3e+03	2694
/SID56/gcgdata/geneseq/geneseq/NA1999.DAT:X176441 +	32.00	96.91	1.3e+03	2694
seq_name: /SID56/gcgdata/geneseq/geneseq/NA1990.DAT:Q06828				
seq_documentation_block:				
ID Q06828 standard; rRNA; 1872 BP.				
XX Q06828;				
AC				
XX				
DT 06-MAR-1991 (first entry)				
XX				
DE Extracellular portion of the human epidermal growth factor receptor 2.				
XX				
KW Human epidermal growth factor receptor 2; HER2; vaccine; cancer; ss.				
XX				
OS Homo sapiens.				
XX				
FH Key	Location/Qualifiers			
FT misc_RNA	64...96			
FT	/tag= a			
FT	/number= 1			
FT	/note= "potential T-cell epitope"			
FT	220...252			
FT	/tag= b			
FT	/number= 2			
FT	/note= "potential T-cell epitope"			
FT	337...390			
FT	/tag= c			
FT	/number= 3			
FT	/note= "potential T-cell epitope"			
FT	559...591			
FT	/tag= d			
FT	/number= 4			
FT	/note= "potential T-cell epitope"			
FT	913...945			
FT	/tag= e			
FT	/number= 5			
FT	/note= "potential T-cell epitope"			
FT	979...1059			
FT	/tag= f			
FT	/number= 6			
FT	/note= "potential T-cell epitope"			
FT	1123...1155			
FT	/tag= g			
FT	/number= 7			
FT	/note= "potential T-cell epitope"			
FT	1192...1224			
FT	/tag= h			
FT	/number= 8			
FT	/note= "potential T-cell epitope"			
FT	1297...1329			
FT	/tag= i			
FT	/number= 9			
FT	/note= "potential T-cell epitope"			
FT	1555...1587			
FT	/tag= j			
FT	/number= 10			
FT	/note= "potential T-cell epitope"			
XX				
XX				
PN WO9014357-A.				
XX				
XX				
PD 29-NOV-1990.				
XX				
XX				
PF 18-MAY-1990; 90WO-US02697.				
XX				
XX				
PR 19-MAY-1989; 89US-0354319.				
XX				
XX				
PA (GETH ) GENENTECH INC.				

PI Hudziak RM, Shepard HM, Ullrich A;  
 XX WPI; 1990-375946/50.  
 DR P-PSDB; R08222.  
 XX HER2 extracellular domain used as vaccine - comprises sequence of  
 PT at least 9 amino acid(s) prep. using expression vector of DNA  
 PT isolated from human epidermal growth factor receptor  
 XX  
 PS Disclosure; Fig 13; 49pp; English.  
 XX  
 CC This claimed sequence terminates 1 bp, pref. 24 bps, upstream  
 CC of the DNA portion encoding the transmembrane domain of the HER2 mol.  
 CC It encodes none of the transmembrane or intracellular portions of  
 CC the HER2 mol. The protein is antigenic in animals.  
 CC The sequence can be used to produce an extracellular portion of  
 CC the HER2 mol. comprising at least 9 amino acids. The product  
 CC opens several possibilities for further research and a broad range  
 CC of potential therapeutic applications, eg. for the treatment of  
 CC mammary gland adenocarcinoma and other cancers.  
 XX  
 SQ Sequence 1872 BP; 377 A; 602 C; 541 G; 352 U; 0 other;

alignment\_scores:  
 Quality: 43.00 Length: 9  
 Ratio: 4.778 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
 US-09-277-074-10 x Q06828 ..

Align seg 1/1 to: Q06828 from: 1 to: 1872

1 LysilePheGlySerLeuAlaPheLeu 9  
 |||||  
 1042 AAGAUUUUGGAGCGGCAUUCUG 1068

seq\_name: /SIDS6/gcgdata/geneseq/geneseq/NA1997.DAT:T72725

seq\_documentation\_block:  
 ID T72725 standard; cdna; 2385 BP.

XX T72725;

XX 17-SEP-1997 (first entry)

XX Her2-GM-CSF immunostimulant fusion protein DNA.

XX Her2-GM-CSF; granulocyte macrophage colony stimulating factor;  
 KW growth factor receptor; oncogene; immunostimulant; cancer;  
 KW therapy; ss.

XX Homo sapiens.

XX Key Location/Qualifiers

FT CDS 11..2359

FT /tag= a  
 FT /product= GM-CSF-Her2 fusion protein

FT mRNA 11..1969

FT /tag= b

FT /product= Her2

FT mRNA 1970..1975

FT /tag= c

FT /product= Leu-Glu linker

FT mRNA 1976..2359

FT /tag= d

FT /product= GM-CSF

XX WO9724438-A1.

PN

XX 10-JUL-1997.

PD

XX

PF 23-DEC-1996; 96WO-US20241.  
 XX  
 PR 28-DEC-1995; 95US-0579823.  
 XX  
 PA (ACTI-) ACTIVATED CELL THERAPY INC.  
 XX  
 XI Laus R, Ruegg CL, Wu H;  
 PI WPI; 1997-363674/33.  
 XX  
 DR P-PSDB; W19764.  
 XX  
 XX Potent APC that activates T-cells to give multivalent cellular  
 PT immune response - can also induce a cytotoxic T-cell response in a  
 PT vertebrate subject  
 PT  
 PS Disclosure; Fig 8; 45pp; English.  
 XX  
 CC A nucleic acid molecule (T72725) codes for a fusion protein  
 CC (W19764) comprising granulocyte-macrophage colony stimulating  
 CC factor (GM-CSF) and Her2, a growth factor receptor that is  
 CC over-expressed in breast and ovarian cancer cells. It was  
 CC prepd. by PCR amplification of Her2 cDNA from a breast cancer  
 CC cell line and fusion to GM-CSF cDNA. Fusion expression vectors can  
 CC be used to transfect mammalian and insect cells. The Her2-GM-CSF  
 CC fusion protein is used to generate anti-Her2 immunity. Tumour  
 CC cells are eliminated by cytotoxic T lymphocytes activated in vivo  
 CC or in vitro by exposure to antigen-presenting cells exposed to the  
 CC fusion protein.  
 XX  
 SQ Sequence 2385 BP; 488 A; 780 C; 677 G; 440 T; 0 other;

alignment\_scores:  
 Quality: 43.00 Length: 9  
 Ratio: 4.778 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
 US-09-277-074-10 x T72725 ..

Align seg 1/1 to: T72725 from: 1 to: 2385

1 LysilePheGlySerLeuAlaPheLeu 9  
 |||||  
 1115 AAGATCTTTGGAGCTGGCAATTCTG 1141

seq\_name: /SIDS6/gcgdata/geneseq/geneseq/NA2000.DAT:250586

seq\_documentation\_block:

ID 250586 standard; DNA; 2871 BP.

XX 250586;

AC

XX 23-MAY-2000 (first entry)

XX DC8scFv-erbB2EC fusion construct containing tetramerisation domain.  
 DE

XX DC8 scFv; single-chain variable fragment; erbB2EC; extracellular domain;  
 KW human; fusion construct; tetramerisation domain; constant domain;  
 KW heteroninibody; multifunctional compound; melanoma; sarcoma;  
 KW immunoglobulin; cytostatic; immunostimulatory; antileukaemia; diagnosis;  
 KW antiproliferative; prevention; treatment; malignant; hematopoietic cell;  
 KW lymphoma; leukaemia; solid tumour; carcinoma; ds.

XX Chimeric - Unidentified.

OS Chimeric - Homo sapiens.

XX Key Location/Qualifiers

FT CDS 10..2865

FT /tag= a

FT /product= "DC8scFv-erbB2EC fusion protein"

FT sig\_peptide 10..66

FT /tag= b

FT misc\_feature /label= Leader\_sequence  
FT 67..390  
FT \*tag= c  
FT /label= DC8scFv\_light\_chain\_variable\_region  
FT 391..435  
FT \*tag= d  
FT /label= Glycine-Serine-linker\_DNA  
FT 436..771  
FT \*tag= e  
FT /label= DC8scFv\_heavy\_chain\_variable\_region  
FT 775..807  
FT \*tag= f  
FT /note= '5' end of human IgG3 upper hinge region  
FT with 3 additional nucleotides"  
FT 808..924  
FT \*tag= g  
FT /label= Human\_p53\_tetramerisation\_domain  
FT 925..945  
FT \*tag= h  
FT /label= Short\_peptide\_linker  
FT 946..2844  
FT \*tag= i  
FT /label= erB2EC\_domain  
FT 2845..2862  
FT \*tag= j  
FT /label= His\_tag  
FT  
FT  
XX WO200006605-A2.  
XX  
XX  
PD 10-FEB-2000.  
XX  
XX 28-JUL-1999; 99WO-EP05416.  
XX  
XX 28-JUL-1998; 98EP-0114082.  
XX  
XX (MICR-) MICROMET GES BIOMEDIZINISCHE FORSCHUNG.  
XX  
XX Kufer P, Dreier T, Baeuerle PA, Borschert K, Zettl F;  
XX WPI: 2000-195265/17.  
XX P-PSDB; Y44993.  
XX  
XX New multifunctional compounds useful for preventing and/or treating  
XX malignant cell growth and for detection and diagnosis -  
XX  
XX Example 9; Fig 49; 166pp; English.  
XX  
XX The patent discloses heteroinibodies which are multifunctional compounds  
XX producible in a mammalian host cell as a secretable and fully functional  
XX heterodimer of two polypeptide chains, where one of the polypeptide  
XX chains comprises a CHI-domain (constant domain of an immunoglobulin  
XX heavy chain) and the other chain comprises CI-domain (constant domain of  
XX an immunoglobulin light chain). The polypeptide chains further comprise,  
XX fused to the constant domains at least two (poly)peptides having  
XX different receptor or ligand functions, where further at least two of the  
XX different (poly)peptides lack an intrinsic affinity for one another and  
XX are linked via the constant domains. The heteroinibodies have  
XX cytostatic, immunostimulatory, antileukemia and antiproliferative  
XX activities. These compounds can be used for diagnosing, preventing and  
XX treating malignant cell growth related to malignancies of hematopoietic  
XX cells e.g. lymphomas and leukaemias, or to solid tumours e.g. carcinomas,  
XX melanomas and sarcomas.  
XX The present sequence is a fusion construct comprising DC8  
XX single-chain Fv (scFv) fragment at the N-terminus, extracellular  
XX domain of human erB2 at the C-terminus and a tetramerisation  
XX domain between them. This construct was prepared to find out whether  
XX an oligomerisation domain characterised in bacterial expression system  
XX is applicable for expression of fully functional and secretable  
XX recombinant protein in mammalian host cells. This tetrameric construct  
XX was not expressed as secretable and fully functional protein  
XX in mammalian cells. Hence general applicability of the tetramerisation  
XX domain for oligomerisation strategies in mammalian cells was ruled out.  
XX

SQ Sequence 2871 BP; 598 A; 868 C; 834 G; 571 T; 0 other;  
  
alignment\_scores:  
Quality: 43.00 Length: 9  
Ratio: 4.778 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000  
  
alignment\_block:  
US-09-277-074-10 x 250586 ..  
  
Align seg 1/1 to: 250586 from: 1 to: 2871  
  
1 LysIlePheGlySerLeuAlaPheLeu 9  
|||||  
1987 ARGATCTTTGGGAGCCTGGCATTTCTG 2013  
  
seq\_name: /SID86/gcgdata/geneseq/geneseqn/NA1996.DAT:T40739  
  
seq\_documentation\_block:  
ID T40739 standard; cDNA; 3768 BP.  
XX  
AC T40739;  
XX  
DT 01-JAN-1997 (first entry)  
XX  
DE HER-2/neu oncogene.  
XX  
KW HER-2/neu; c-erbB1; p185; oncogene; tyrosine protein kinase;  
KW breast cancer; ovary cancer; colon cancer; lung cancer;  
KW prostate cancer; genetic immunisation; tumour; vaccine; vector;  
KW ss.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT 1..3765  
FT CDS /tag= b  
FT /note= "nucleotides 2026-3765 (claim 1) code for  
FT HER-2/neu intracellular domain"  
XX  
PN WO9630514-A1.  
XX  
PD 03-OCT-1996.  
XX  
PF 28-MAR-1996; 96WO-US01689.  
XX  
PR 31-MAR-1995; 95US-0414417.  
XX (UNIW ) UNIV WASHINGTON.  
XX Cheever MA, Disis ML;  
XX WPI; 1996-455361/45.  
XX P-PSDB; W01111.  
XX  
XX DNA encoding HER-2/neu poly:peptide(s) - used for prevention or  
XX treatment of malignancies with which the HER-2/neu oncogene is  
XX associated  
XX  
PS Claim 1; Page 49-56; 71pp; English.  
XX  
XX Human HER-2/neu oncogene cDNA (T40739) codes for HER-2/enu (p185 or  
XX c-erbB2) protein (W01111). The oncogene is overexpressed in various  
XX cancers, including breast, ovarian, colon, lung and prostate, and  
XX appears to induce malignancies through quantitative mechanisms that  
XX result from increased or deregulated expression of an essentially  
XX normal gene product. Nucleotides 2026-3765 of the cDNA sequence  
XX code for the intracellular domain (Lys676-Vall255) of the HER-2/neu  
XX protein, which is useful for immunisation against malignancy.  
XX Nucleic acids can be used to direct expression of the intracellular  
XX domain in transformed host cells, or are used, alone or in a viral  
XX vector, for genetic immunisation of an animal.

XX Sequence 3768 BP; 759 A; 1171 C; 1119 G; 719 T; 0 other;

alignment\_scores:  
Quality: 43.00 Length: 9  
Ratio: 4.778 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-09-277-074-10 x T40739 ..

Align seg 1/1 to: T40739 from: 1 to: 3768

1 LysllePheGlySerLeuAlaPheLeu 9  
|||||  
1105 AAGATCTTTGGAGCGCTGGCATTCTG 1131

seq\_name: /SIDS6/gcgdata/geneseq/NA1999.DAT:X01912

seq\_documentation\_block:  
ID X01912 standard; DNA; 3768 BP.

XX AC X01912;

XX 21-APR-1999 (first entry)

XX Human HER-2/neu oncogene DNA.

XX HER-2/neu; oncogene; immune response; T cell; B cell; immunisation;  
XX malignancy; treatment; tumour; ss.

XX Homo sapiens.

XX Key Location/Qualifiers  
XX CDS 1..3768  
XX /\*tag= a  
XX /product= "HER-2/neu"  
XX /note= "oncogene"  
XX 2026..3765  
XX misc\_feature /\*tag= b  
XX /note= "region which elicits immune response"

XX US5869445-A.

XX 09-FEB-1999.

XX 01-APR-1996; 96US-0625101.

XX 01-APR-1996; 96US-0625101.

XX 17-MAR-1993; 93US-0033644.

XX 12-AUG-1993; 93US-0106112.

XX 31-MAR-1995; 95US-0414417.

XX (UNIW ) UNIV WASHINGTON.

XX Cheever MA, Disis ML;

XX WPI; 1999-152835/13.

XX P-PSDB; W92406.

XX Use of HER-2/neu polypeptides - for eliciting an immune response to  
XX an HER-2/neu associated malignancy, particularly for treating or  
XX preventing tumours

XX Claim 1a; Column 23-32; 26pp; English.

XX This sequence encodes the human HER-2/neu oncogene protein. A fragment  
XX of this protein is used in a method for eliciting or enhancing an immune  
XX response to HER-2/neu protein. The polypeptide can stimulate T cells and  
XX B cells to produce an immune response to the HER-2/neu protein. The  
XX method can be used for immunisation against a malignancy in which the  
XX HER-2/neu oncogene is associated and in the treatment of an existing

CC tumour, or to prevent tumour occurrence or reoccurrence.  
XX Sequence 3768 BP; 759 A; 1171 C; 1119 G; 719 T; 0 other;

alignment\_scores:  
Quality: 43.00 Length: 9  
Ratio: 4.778 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-09-277-074-10 x X01912 ..

Align seg 1/1 to: X01912 from: 1 to: 3768

1 LysllePheGlySerLeuAlaPheLeu 9  
|||||  
1105 AAGATCTTTGGAGCGCTGGCATTCTG 1131

seq\_name: /SIDS6/gcgdata/geneseq/NA2000.DAT:A09455

seq\_documentation\_block:  
ID A09455 standard; DNA; 3768 BP.

XX AC A09455;

XX 10-AUG-2000 (first entry)

XX Human heregulin 2 (Her2) coding sequence.

XX Heregulin 2; Her2; vaccination; cytotoxic T-lymphocyte immunity;  
XX self-protein; cell-associated peptide antigen; foreign epitope;  
XX cancer; breast cancer; prostate cancer; ss.

XX Homo sapiens.

XX WO200020027-A2.

XX 13-APR-2000.

XX 05-OCT-1999; 99WO-DK00525.

XX 05-OCT-1998; 98DK-0001261.

XX 20-OCT-1998; 98US-0105011.

XX (NEBI-) M & E BIOTECH AS.

XX Steinaa L, Mouritsen S, Nielsen KG, Haaning J, Leach D, Dalum I;  
XX Gautam A, Birk P, Karlsson G;

XX WPI; 2000-349917/30.

XX P-PSDB; Y92620.

XX Inducing immune responses to weakly immunogenic, tumor associated  
XX peptide antigens for the treatment of breast and prostate cancer

XX Claim 62; Page 187-193; 220pp; English.

XX The claims detail a method for inducing immune responses against weakly  
XX immunogenic cell-associated peptide antigens (PA) such as those  
XX associated with cancers (i.e. self-proteins), for example, human  
XX prostate specific membrane antigen (PSM), heregulin 2 (Her2) and/or  
XX fibroblast growth factor 8b (FGF8b). The method comprises effecting  
XX simultaneous presentation by antigen producing cells (APCs) of the  
XX animals immune system of: (1) at least 1 CTL (cytotoxic T-lymphocyte)  
XX group derived from the PA and/or at least 1 B-cell group derived from the  
XX cell-associated PA; and (2) at least 1 first T helper cell group which is  
XX foreign to the animal. Analogues of human PSM, human Her2 and  
XX human/murine FGF8b comprising a substantial part of all known and  
XX predicted CTL and B-cell epitopes of the respective PA and including at  
XX least one foreign T helper epitope are also claimed. The method is used  
XX to treat prostate, prostate/breast or breast cancer when the PA is human  
XX PSM, FGF8b and Her2, respectively.



XX SQ Sequence 3768 BP; 758 A; 1170 C; 1121 G; 719 T; 0 other;

alignment\_scores:  
Quality: 43.00 Length: 9  
Ratio: 4.778 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-09-277-074-10 x A09455 ..

Align seg 1/1 to: A09455 from: 1 to: 3768

1 LysilePheGlySerLeuAlaPheLeu 9  
|||||  
1105 AAGATCTTTGGAGCCTGGCATTTCTG 1131

seq\_name: /SID56/gcgdata/geneseq/geneseq/NA1995.DAT:T01590

seq\_documentation\_block:  
ID T01590 standard; DNA; 3955 BP.

XX AC T01590;  
XX DT 21-APR-1996 (first entry)  
XX DE Rat neu promoter.

XX KW Her-2/neu; c-erbB-2; oncogene; Erb-B2 promoter; DNA binding protein;  
XX KW HPBF; Erb-B2 promoter binding protein; tumour enhancer factor;  
XX KW breast cancer diagnosis; prognosis; antisense oligonucleotide;  
XX KW retro virus vector; gene therapy vector; ss.  
XX OS Rattus rattus.  
XX PN W09528485-A1.  
XX PD 26-OCT-1995.  
XX PF 19-APR-1995; 95WO-US04953.  
XX PR 19-APR-1994; 94US-0229515.  
XX PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
XX PI Raziuddin F, Sarkar FH;  
XX DR WPI; 1995-373800/48.  
XX PT New purified protein binding to the ERBB2 gene promoter - to induce  
XX PT cell proliferation, diagnostic of breast cancer, also related  
XX PT antibodies, nucleic acid, assays and methods for screening  
XX PT inhibitors.  
XX PS Disclosure; Page 57-59; 69pp; English.

XX CC The sequence represents one of the promoters of the rat neu gene;  
XX CC a second rat neu gene promoter is in T01589. These promoter are  
XX CC compared with the human Erb-B2 promoters (T01587-88) and the mouse  
XX CC neu promoter (T01591). The human Erb-B2 gene is one of the  
XX CC primary genes responsible for the transition of normal breast  
XX CC epithelial cells towards carcinoma in situ and the subsequent  
XX CC development of invasive and metastatic cancer. The Erb-B2;  
XX CC promoter-binding protein, HPBF (see R77093-94), induces cell  
XX CC division on binding to the promoter. In a method for greater  
XX CC success in early identification and treatment of breast cancer,  
XX CC the initiation step for Erb-B2 gene activity is identified. This  
XX CC method involves determining the presence of HPBF in a biopsy from  
XX CC the subject, where the presence of HPBF (relative to its absence  
XX CC in a normal control) indicates the presence of cancer and a  
XX CC decreased chance of long-term survival. Binding of HPBF to the  
XX CC promoter can be inhibited using antisense oligonucleotides or a

CC non-genomic nucleic acid that binds to HPBF; these oligos can be  
CC expressed from retro virus or other gene therapy vectors.

XX SQ Sequence 3955 BP; 842 A; 1147 C; 1136 G; 830 T; 0 other;

alignment\_scores:  
Quality: 43.00 Length: 9  
Ratio: 4.778 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-09-277-074-10 x T01590 ..

Align seg 1/1 to: T01590 from: 1 to: 3955

1 LysilePheGlySerLeuAlaPheLeu 9  
|||||  
1133 AAGATCTTTGGAGCCTGGCATTTTG 1159

seq\_name: /SID56/gcgdata/geneseq/geneseq/NA2000.DAT:A14812

seq\_documentation\_block:  
ID A14812 standard; cDNA; 4472 BP.

XX AC A14812;  
XX DT 08-AUG-2000 (first entry)  
XX DE cDNA encoding the SPLICE erBB-2 receptor protein.

XX KW SPLICE erBB-2 receptor protein; cell transformation disorder; cancer;  
XX KW tumor cell proliferation; tissue degeneration; arthropathy;  
XX KW bone resorption; inflammatory disease; degenerative disorder;  
XX KW wound healing; ss.  
XX OS Homo sapiens.

XX FH Key Location/Qualifiers  
XX FT CDS 175..3942  
XX FT /\*tag= a  
XX FT /product= "SPLICE erBB-2 protein"

XX PN W0200020579-A1.  
XX PD 13-APR-2000.  
XX PF 01-OCT-1999; 99WO-CA00912.  
XX PR 02-OCT-1998; 98US-0165192.  
XX PA (UYMC-) UNIV MCMASTER.  
XX PI Muller WJ, Siegel PM;  
XX PS WPI; 2000-303768/26.  
XX DR P-PSDB; Y84780.

XX PT Nucleic acid encoding an erBB 2 receptor protein designated SPLICE  
XX PT erBB-2, inhibitors of the protein are useful for treatment of cancer -  
XX PS Claim 4; Fig 1; 60pp; English.

XX CC The present sequence encodes a SPLICE erBB-2 receptor protein. The  
XX CC protein has an in-frame deletion of 16 amino acids, 2 of which are  
XX CC conserved cysteine residues, compared to the unspliced protein. The  
XX CC erBB-2 polynucleotide is used to construct probes for detecting  
XX CC disorders of cell transformation such as cancer. Antibodies to the  
XX CC protein may be used to detect SPLICE erBB-2 in a sample. Agents  
XX CC (e.g. antisense oligonucleotides) which inhibit the expression of  
XX CC SPLICE erBB-2 are useful for reducing tumor cell proliferation and  
XX CC treating cancer. Substances which stimulate SPLICE erBB-2 are useful  
XX CC for treating conditions involving damaged cells including conditions

CC in which degeneration of tissue occurs, such as arthropathy, bone  
 CC resorption, inflammatory diseases, degenerative disorders of the  
 CC central nervous system and wound healing.

XX Sequence 4472 BP; 902 A; 1383 C; 1328 G; 859 T; 0 other;

alignment\_scores:  
 Quality: 43.00 Length: 9  
 Ratio: 4.778 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
 US-09-277-074-10 x A14812 ..

Align seg 1/1 to: A14812 from: 1 to: 4472

1 LysilePheGlySerLeuAlaPheLeu 9  
 |||||  
 1279 AGATCTTTGGAGCCTGGCATTTCTG 1305

seq\_name: /SIDS6/gcgdata/geneseq/geneseqn/NA1999.DAT:231071

seq\_documentation\_block:  
 ID 231071 standard; DNA; 4473 BP.

XX AC 231071;  
 XX DT 17-JAN-2000 (first entry)  
 XX DE HER-2 nucleic acid sequence.

XX KW HER-2; c-neu; ErbB2; transmembrane receptor; tyrosine kinase activity;  
 KW epidermal growth factor receptor; EGFR; HER-1; cancer; breast cancer;  
 KW ovarian cancer; gastric cancer; antisense oligonucleotide; expression;  
 KW hyperproliferative disease; ss.

XX OS Homo sapiens.  
 XX PN WO9948906-A1.  
 XX PD 30-SEP-1999.

XX PF 25-MAR-1999; 99WO-US06492.  
 XX PR 26-MAR-1998; 98US-0048804.  
 XX PA (ISIS-) ISIS PHARM INC.  
 XX PA (PENN-) PENN STATE RES FOUND.

XX PI Bennett CF, Lipton A, Witters LM;  
 XX DR WPI; 1999-610749/52.  
 XX PT New antisense sequences used to treat hyperproliferative conditions,  
 XX especially cancer -

XX PS Examples; Page 38-39; 44pp; English.  
 XX This is the human HER-2 polynucleotide sequence. The HER-2 gene also  
 CC called c-neu and ErbB2, encodes a transmembrane receptor, with tyrosine  
 CC kinase activity. HER-2 is related to the epidermal growth factor receptor  
 CC (EGFR or HER-1). Aberrant HER-2 expression is present in a wide number  
 CC of cancers, especially breast, ovarian and gastric cancers. This sequence  
 CC is used in the invention to design 12-25 nucleotide oligonucleotides that  
 CC decrease the expression of human HER-2. The oligonucleotides of the  
 CC invention (231057-231070) can also be used for modulating the expression  
 CC of human epidermal growth factor receptor. The oligonucleotides are used  
 CC to treat diseases or conditions associated with HER-2, particularly  
 CC hyperproliferative diseases such as cancer.

XX Sequence 4473 BP; 902 A; 1383 C; 1329 G; 859 T; 0 other;

alignment\_scores:  
 Quality: 43.00 Length: 9  
 Ratio: 4.778 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
 US-09-277-074-10 x Z31071 ..

Align seg 1/1 to: Z31071 from: 1 to: 4473

1 LysilePheGlySerLeuAlaPheLeu 9  
 |||||  
 1279 AGATCTTTGGAGCCTGGCATTTCTG 1305

seq\_name: /SIDS6/gcgdata/geneseq/geneseqn/NA1995.DAT:T01585

seq\_documentation\_block:  
 ID T01585 standard; DNA; 4530 BP.

XX AC T01585;  
 XX DT 20-APR-1996 (first entry)  
 XX DE Her-2/neu (ERBB2/c-erbB-2) gene sequence.

XX KW Her-2/neu; Erb-B2; c-erbB-2; oncogene; DNA binding protein; HPBF;  
 KW Erb-B2 promoter binding protein; tumour enhancer factor;  
 KW breast cancer diagnosis; prognosis; antisense oligonucleotide;  
 KW retro virus vector; gene therapy vector; ss.

XX OS Homo sapiens.  
 XX PN WO9528485-A1.  
 XX PD 26-OCT-1995.

XX PF 19-APR-1995; 95WO-US04953.  
 XX PR 19-APR-1994; 94US-0229515.  
 XX PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.

XX PI Raziuddin F, Sarkar FH;  
 XX DR WPI; 1995-373800/48.  
 XX PT New purified protein binding to the ERBB2 gene promoter - to induce  
 XX cell proliferation, diagnostic of breast cancer, also related  
 XX PT antibodies, nucleic acid, assays and methods for screening  
 XX PT inhibitors.

XX PS Disclosure; Page 52-54; 69pp; English.  
 XX The Erb-B2 gene is one of the primary genes responsible for the  
 CC transition of normal breast epithelial cells towards carcinoma in  
 CC situ and the subsequent development of invasive and metastatic  
 CC cancer. HPBF (see R77093-94), the Erb-B2 promoter binding protein,  
 CC induces cell division on binding to the promoter. In a method for  
 CC greater success in early identification and treatment of breast  
 CC cancer, the initiation step for Erb-B2 gene activity is identified.  
 CC This method involves determining the presence of HPBF in a biopsy  
 CC from the subject, where the presence of HPBF (relative to its  
 CC absence in a normal control) indicates the presence of cancer and  
 CC a decreased chance of long-term survival. Binding of HPBF to the  
 CC promoter can be inhibited using antisense oligonucleotides or a  
 CC non-genomic nucleic acid that binds to HPBF; these oligos can be  
 CC expressed from retro virus or other gene therapy vectors.

XX Sequence 4530 BP; 922 A; 1382 C; 1346 G; 880 T; 0 other;

alignment\_scores:

Quality: 43.00 Length: 9  
Ratio: 4.778 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-09-277-074-10 x T01585 ..

Align seg 1/1 to: T01585 from: 1 to: 4530

1 LysilePheGlySerLeuAlaPheLeu 9

|||||  
1255 AAGATCTTTGGAGCCTGGCATTCTG 1281

seq\_name: /SID56/gcgdata/geneseq/geneseqn/NA197.DAT:T71253

## seq\_documentation\_block:

ID T71253 standard; DNA; 4530 BP.

XX

AC

XX

DT

XX

DE Human HER2 gene.

KW

KW

KW

KW

KW

XX

XX

XX

PN

XX

XX

PD

XX

XX

PF

XX

PR

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

HER2; cognate transgene; human; tyrosine kinase-type receptor; lymphoma;  
cellular immunogen; cancer; self-determinant immunoreactivity;  
cancer vaccination; breast carcinoma; colon carcinoma; immunotherapy;  
proto-oncogene; ss.

Homo sapiens.

WO9725860-A1.

24-JUL-1997.

13-JAN-1997; 97WO-US00582.

19-JAN-1996; 96US-0010262.

(UYAL-) UNIV ALLEGHENY HEALTH SCI.

England JM, Halpern MS;

WPI; 1997-384993/35.

Proto-oncogene immunogen - used in vaccine for the prevention and  
treatment of cancer

Disclosure; Page 56-58; 81pp; English.

This sequence represents the human HER2 cognate transgene (CTG).  
Deletion of amino acids 1-731 of the encoded protein renders the CTG  
non-transforming. HER2 is a tyrosine kinase-type receptor. This sequence  
can be used in the cellular immunogen of the invention. The cellular  
immunogen of the invention is for immunising against the product of a  
target proto-oncogene, over-expression of which is associated with  
cancer, comprises host cells transfected with a construct containing at  
least one transgene related to the proto-oncogene and driven by a strong  
promoter. The product of the transgene induces immunoreactivity to host  
self-determinants on the product of proto-oncogene. The cellular  
immunogens are used for protective vaccination against cancer (e.g.  
carcinoma of breast or colon, or various lymphomas) and for immunotherapy  
of cancer. Use of the immunogen eliminates the need to isolate  
immunogenic, HLA host-matched peptides. The method is not based on immune  
recognition of a determinant defined by a cancer-specific mutation, and  
generates a systemic (anti-metastatic) response.

Sequence 4530 BP; 922 A; 1382 C; 1346 G; 880 T; 0 other;

## alignment\_scores:

Quality: 43.00 Length: 9  
Ratio: 4.778 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-09-277-074-10 x T71253 ..

Align seg 1/1 to: T71253 from: 1 to: 4530

1 LysilePheGlySerLeuAlaPheLeu 9

|||||  
1255 AAGATCTTTGGAGCCTGGCATTCTG 1281

seq\_name: /SID56/gcgdata/geneseq/geneseqn/NA2000.DAT:Z60815

## seq\_documentation\_block:

ID Z60815 standard; DNA; 4530 BP.

XX

AC

XX

DT

XX

DE

XX

KW

KW

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Nucleotide sequence of a cognate transgene of c-neu.

Cognate transgene; CTG; tumorigenic; cellular immunogen; immunisation;  
proto-oncogene; malignancy; allogenic cell; vaccine; cancer; ss.

Homo sapiens.

WO200004927-A1.

03-FEB-2000.

08-JUL-1999; 99WO-US15594.

24-JUL-1998; 98US-0093965.

(UYAL-) UNIV ALLEGHENY HEALTH SCI.

(HALP/) HALPERN M S.

(ENGL/) ENGLAND J M.

Halpern MS, England JM;

WPI; 2000-182543/16.

Cellular immunogens comprising allogenic donor cells transfected with a  
construct comprising a proto-oncogene cognate, useful as cancer  
vaccines

Disclosure; Page 66-68; 77pp; English.

The present sequence represents a cognate transgene (CTG) which is  
rendered non-tumorigenic by deletion of amino acids 1-731. The CTG  
is used in the course of the invention. The specification describes a  
cellular immunogen for immunizing a host against the effects of the  
product of a target proto-oncogene which is associated with a  
malignancy. The cellular immunogen comprises allogenic cells  
transfected with transgene construct comprising a transgene cognate  
to target proto-oncogene and a strong promoter. The cellular immunogen  
is useful for vaccinating a host against cancer by inserting the  
transgene construct into the body of the host for the expression of  
the transgene. The method of the invention is designed to target  
mutation-driven non-self determinants. The cellular immunogens induce  
reactivity for self-determinants in the over expressed product of tumour  
associated and over expressed proto-oncogenes.

Sequence 4530 BP; 922 A; 1382 C; 1346 G; 880 T; 0 other;

## alignment\_scores:

Quality: 43.00 Length: 9  
Ratio: 4.778 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

```

US-09-277-074-10 x 260815
Align seg 1/1 to: 260815 from: 1 to: 4530
1 LysIlePheGlySerLeuAlaPheLeu 9
|||||
1255 AAGATCTTGGGAGCTGCCTTCG 1281

seq_name: /SIDS6/gcgdata/geneseq/geneseq/NA1998.DAT:Q46083
seq_documentation_block:
ID Q46083 standard; CDNA; 4299 BP.
XX AC Q46083;
XX DT 07-FEB-1994 (first entry)
XX DE Sequence encoding a c-erbB-2 tumour antigen.
XX KW Tumour antigen; c-erbB-2; glycoprotein; ss.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX FT CDS 1..4299
XX FT /*tag= a
XX PN W09316185-A.
XX PD 19-AUG-1993.
XX PF 05-FEB-1993; 93WO-US01055.
XX PR 06-FEB-1992; 92US-0831967.
XX PA (CETUS ONCOLOGY CORP.
XX PA (CREA-) CREATIVE BIOMOLECULES INC.
XX PI Houston LL, Huston JS, Oppermann H, Ring DB;
XX WPI; 1993-272889/34.
XX DR P-PSDB; R39568.
XX PT New single chain Fv polypeptide binding to C-erbB-2 tumour
XX PT antigen - for imaging or treating breast or ovarian cancer etc.
XX PS Disclosure; pages 48-54; 87pp; English.
XX CC c-erbB-2 refers to a protein antigen expressed on the surface of
XX CC tumour cells, such as breast and ovarian tumour cells, which is an
XX CC approx. 200,000 mol. wt. acidic glycoprotein having an isoelectric
XX CC pt. of about 5.3 (see Q46083, R39568).
XX SQ Sequence 4299 BP; 882 A; 1307 C; 1266 G; 844 T; 0 other;

alignment_scores:
Quality: 40.00 Length: 9
Ratio: 4.444 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 88.889

alignment_block:
US-09-277-074-10 x Q46083
Align seg 1/1 to: Q46083 from: 1 to: 4299
1 LysIlePheGlySerLeuAlaPheLeu 9
|||||
1105 AAGATCTTGGGAGCTGCCTTCG 1131

seq_name: /SIDS6/gcgdata/geneseq/geneseq/NA1998.DAT:V65286
seq_documentation_block:

```

```

ID V65286 standard; DNA; 969 BP.
XX AC V65286;
XX DT 24-DEC-1998 (first entry)
XX DE DNA encoding a S. pneumoniae glycosyl transferase.
XX KW Streptococcus pneumoniae protein; recombinant; gene expression;
XX KW DNA chip; virulence; antibody; infection; detection; treatment; ss.
XX OS Streptococcus pneumoniae.
XX PN W09826072-A1.
XX PD 18-JUN-1998.
XX PF 09-DEC-1997; 97WO-US22578.
XX PR 13-DEC-1996; 96US-0036281.
XX PA (ELIL ) LILLY & CO ELI.
XX PI Baltz RH, Burgett SG, Dehoff BS, Hoskins JA, Jaskunas SR;
XX PI Mills BJ, Norris FH, Peery RB, Rostock PK, Rostock PR;
XX PI Skatrud PL, Smith MC, Solenberg PJ, Treadway PJ;
XX PI Young Bellido ML;
XX WPI; 1998-348529/30.
XX DR P-PSDB; W80728.
XX PT Streptococcus pneumoniae nucleic acid sequences - used in DNA chips
XX PT for evaluating gene expression, and identification of virulence
XX PS genes
XX PS Claim 1; Page 151; 333pp; English.
XX CC This DNA sequence encodes a Streptococcus pneumoniae glycosyl
XX CC transferase. The invention provides DNA sequences (V65201 to V65304) from
XX CC the Streptococcus pneumoniae genome and corresponding protein sequences
XX CC (W80605 to W80728). A recombinant host containing a vector comprising any
XX CC of the above nucleic acids can be used for the recombinant expression of
XX CC the protein sequences. The invention also provides a DNA chip having
XX CC arrayed on it at least 15 base pair fragment of any one or more of these
XX CC DNA sequences. The DNA chip can be used methods for evaluating gene
XX CC expression in S. pneumoniae and for identifying virulence genes in
XX CC S. pneumoniae. Antibodies that selectively bind to the above proteins or
XX CC peptide fragments can be used to treat S. pneumoniae infection. The
XX CC antibodies can also be used to detect S. pneumoniae cells.
XX SQ Sequence 969 BP; 370 A; 97 C; 172 G; 330 T; 0 other;

alignment_scores:
Quality: 37.00 Length: 9
Ratio: 4.111 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 88.889

alignment_block:
US-09-277-074-10 x V65286/rev
Align seg 1/1 to reverse of: V65286 from: 1 to: 969
1 LysIlePheGlySerLeuAlaPheLeu 9
|||||
740 AAAATTTTCTCTCTGCTTCTTG 714

seq_name: /SIDS6/gcgdata/geneseq/geneseq/NA1998.DAT:V52288
seq_documentation_block:
ID V52288 standard; DNA; 6474 BP.
XX AC V52288;

```

XX	23-OCT-1998	(first entry)
XX	DT	
XX	DE	Streptococcus pneumoniae genome fragment SEQ ID NO:155.
XX	DE	Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;
XX	KW	computer readable medium; vaccine; pharmaceutical composition; ds.
XX	XX	
OS	Streptococcus pneumoniae.	
XX	WO9818931-A2.	
PN	07-MAY-1998.	
XX		
PD	30-OCT-1997;	97WO-US19588.
PF		
PR	31-OCT-1996;	96US-0029960.
XX		
XX	(HUMA-)	HUMAN GENOME SCI INC.
PA		
XX	Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M;	
PI	Kunsch CA, Rosen CA;	
PI		
XX	WPI; 1998-272225/24.	
DR		
XX	Computer-readable medium with recorded Streptococcus pneumoniae	
XX	polynucleotide sequences - useful in diagnostic kits and assays, and	
PT	pharmaceutical compositions and vaccines for Streptococcus	
PT	pneumoniae	
XX		
PS	Claim 1; Page 1021-1025; 1409pp; English.	
XX		
CC	The present invention describes a computer readable medium which has	
CC	the nucleotide sequences SEQ ID NO:1 to 391 (V52134 to V52524) recorded	
CC	on it, or a representative fragment or a sequence at least 95% identical	
CC	to SEQ ID NO: 1 to 391. The nucleotide sequences depicted in SEQ ID NO:1	
CC	to 391 (V52134 to V52524) are genomic fragments from Streptococcus	
CC	pneumoniae. The present invention also describes an isolated nucleic acid	
CC	molecule encoding a homologue of any of the fragments of the S.pneumoniae	
CC	genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced	
CC	by a process comprising: (a) screening a genomic DNA library using as a	
CC	probe a target sequence defined by any of the sequences in SEQ ID NO:1	
CC	to 391, identifying members of the library which contain sequences	
CC	that hybridise to the target sequence and isolating the nucleic acid	
CC	molecules from the members; or (b) isolating mRNA, DNA or cDNA produced	
CC	from an organism, amplifying nucleic acid molecules whose nucleotide	
CC	sequence is homologous to amplification primers derived from the	
CC	fragment of the S. pneumoniae genome to prime the amplification and	
CC	isolating the amplified sequences. The computer readable medium can be	
CC	used in a computer-based system for identifying fragments of the	
CC	S. pneumoniae genome of commercial importance, or expression modulating	
CC	fragments of the S. pneumoniae genome. Products of expression modulating	
CC	invention can be used in diagnosis kits and assays, and pharmaceutical	
CC	compositions and vaccines for S. pneumoniae.	
XX		
XX	Sequence 6474 BP; 2228 A; 1182 C; 717 G; 2347 T; 0 other;	

alignment_scores:		
Quality:	37.00	Length: 9
Ratio:	4.111	Gaps: 0
Percent Similarity:	100.000	Percent Identity: 88.889

```

alignment_block:
US-09-277-074-10 x V52288 ..
Align seg 1/1 to: V52288 from: 1 to: 6474

1 LysllePheGlySerLeuAlaPheLeu 9
|||||:::|||||
4583 AAAATTTTTTCTCTGCGCTCTTG 4609

```



OM of: US-09-277-074-10 to: Issued\_Patents\_NA.\* out\_format : pfs  
 Date: Nov 15, 2000 10:27 PM  
 About: Results were produced by the GenCore software, version 4.5,  
 Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL-frame-pfn.model  
 -Q/cgn2\_1/USPTO\_spool/US09277074/runat\_14112000\_120309\_22628/app\_query.fasta\_1.67  
 -DB-Issued\_Patents\_NA -QFMT=Fastap -SUFFIX=rni -GAPOP=12.000  
 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000  
 -GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500  
 -GAPOP=6.000 -GAPEXT=7.000 -XGAPOP=10.000 -XGAPEXT=0.500  
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62  
 -TRANS-human4.cdi -LIST=45 -DOCLALIGN=200 -THR\_SCORE=pcpt  
 -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs  
 -NORM=ext -MINLEN=0 -MAXLEN=2000000000  
 -USER=US09277074 -CGCN1\_1\_43 -NCPU=6 -ICPU=3 -LONGLOG -NO\_XLPPY  
 -WAIT -THREADS=1

Search information block:

Query: US-09-277-074-10  
 Query length: 9  
 Database: Issued\_Patents\_NA.\*  
 Database sequences: 262060  
 Database length: 75620727  
 Search time (sec): 79.010000

score\_list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
/cgn2_6/ptodata/1/ina/6/COMB.seq:US-08-422-108-2 +	43.00	148.82	0.9104	1872	1
/cgn2_6/ptodata/1/ina/5B.COMB.seq:US-09-146-283-3 +	43.00	146.61	1.21	2385	1
/cgn2_6/ptodata/1/ina/6/COMB.seq:US-08-579-822A-3 +	43.00	146.61	1.21	2385	1
/cgn2_6/ptodata/1/ina/5C.COMB.seq:US-08-625-101-1 +	43.00	142.44	2.07	3768	1
/cgn2_6/ptodata/1/ina/5C.COMB.seq:US-08-356-786-1 +	43.00	142.44	2.07	3768	1
/cgn2_6/ptodata/1/ina/5A.COMB.seq:US-08-229-515A-14 +	43.00	141.99	2.19	3955	1
/cgn2_6/ptodata/1/ina/5A.COMB.seq:US-08-645-865-14 +	43.00	141.99	2.19	3955	1
/cgn2_6/ptodata/1/ina/5B.COMB.seq:US-09-048-804-1 +	43.00	140.87	2.52	4473	1
/cgn2_6/ptodata/1/ina/5A.COMB.seq:US-08-229-515A-9 +	43.00	140.75	2.56	4530	1
/cgn2_6/ptodata/1/ina/5A.COMB.seq:US-08-645-865-9 +	43.00	140.75	2.56	4530	1
/cgn2_6/ptodata/1/ina/5A.COMB.seq:US-08-832-883-50 +	33.00	109.76	136.46	1618	1
/cgn2_6/ptodata/1/ina/5C.COMB.seq:US-08-832-877-50 +	33.00	109.76	136.46	1618	1
/cgn2_6/ptodata/1/ina/6/COMB.seq:US-08-375-703-5 +	32.00	101.07	416.19	2694	1
/cgn2_6/ptodata/1/ina/5B.COMB.seq:US-08-426-627-3 -	32.00	99.81	489.07	3092	1
/cgn2_6/ptodata/1/ina/5C.COMB.seq:US-08-477-396A-3 -	32.00	99.71	495.37	3126	1
/cgn2_6/ptodata/1/ina/5B.COMB.seq:US-08-426-627-5 -	32.00	99.35	519.02	3253	1
/cgn2_6/ptodata/1/ina/5C.COMB.seq:US-08-835-231-17 -	31.00	123.77	22.63	144	1
/cgn2_6/ptodata/1/ina/5B.COMB.seq:US-08-253-155A-2 -	31.00	108.18	167.09	794	1
/cgn2_6/ptodata/1/ina/backfiles1.seq:5472844-3 +	31.00	107.12	191.49	892	1
/cgn2_6/ptodata/1/ina/6/COMB.seq:US-08-072-064-9 +	31.00	105.59	233.08	1055	1
/cgn2_6/ptodata/1/ina/6/COMB.seq:US-08-072-064-9 +	31.00	99.89	484.29	1970	1
/cgn2_6/ptodata/1/ina/PCTUS.COMB.seq:PCT-US92-08558-3 -	31.00	99.48	510.01	2059	1
/cgn2_6/ptodata/1/ina/6/COMB.seq:US-08-072-064-2 -	31.00	99.45	512.04	2066	1
/cgn2_6/ptodata/1/ina/6/COMB.seq:US-08-072-064-3 -	31.00	99.45	512.04	2066	1
/cgn2_6/ptodata/1/ina/6/COMB.seq:US-08-072-064-5 -	31.00	99.45	512.04	2066	1
/cgn2_6/ptodata/1/ina/6/COMB.seq:US-08-072-064-7 -	31.00	99.45	512.04	2066	1
/cgn2_6/ptodata/1/ina/5D.COMB.seq:US-08-738-172-1 -	31.00	96.24	772.74	2936	1
/cgn2_6/ptodata/1/ina/5D.COMB.seq:US-08-738-172-2 -	31.00	92.44	1.3e+03	4454	1
/cgn2_6/ptodata/1/ina/5D.COMB.seq:US-08-738-172-1 -	31.00	89.42	1.9e+03	6201	1
/cgn2_6/ptodata/1/ina/5B.COMB.seq:US-08-325-426B-1 -	31.00	84.42	3.5e+03	10718	1
/cgn2_6/ptodata/1/ina/5B.COMB.seq:US-08-614-770A-1 -	31.00	70.49	2.1e+04	49272	1
/cgn2_6/ptodata/1/ina/6/COMB.seq:US-08-947-823-1 +	31.00	70.01	2.2e+04	51952	1
/cgn2_6/ptodata/1/ina/5B.COMB.seq:US-08-480-784-17 +	30.00	107.97	171.65	522	1
/cgn2_6/ptodata/1/ina/5B.COMB.seq:US-08-483-553-17 +	30.00	107.97	171.65	522	1
/cgn2_6/ptodata/1/ina/5B.COMB.seq:US-08-487-002-17 +	30.00	107.97	171.65	522	1
/cgn2_6/ptodata/1/ina/5B.COMB.seq:US-08-483-554B-17 +	30.00	107.97	171.65	522	1
/cgn2_6/ptodata/1/ina/5B.COMB.seq:US-08-488-011B-17 +	30.00	107.97	171.65	522	1
/cgn2_6/ptodata/1/ina/PCTUS.COMB.seq:PCT-US95-10202-17 +	30.00	107.97	171.65	522	1
/cgn2_6/ptodata/1/ina/PCTUS.COMB.seq:PCT-US95-10203-17 +	30.00	107.97	171.65	522	1

/cgn2\_6/ptodata/1/ina/PCTUS.COMB.seq:PCT-US95-10220-17 + 30.00 107.97 171.65 576  
 /cgn2\_6/ptodata/1/ina/5A.COMB.seq:US-08-086-428B-34 - 30.00 107.07 192.62 576  
 /cgn2\_6/ptodata/1/ina/5C.COMB.seq:US-08-468-570-34 - 30.00 107.07 192.62 576  
 /cgn2\_6/ptodata/1/ina/5C.COMB.seq:US-08-290-665A-34 - 30.00 107.07 192.62 576  
 /cgn2\_6/ptodata/1/ina/PCTUS.COMB.seq:PCT-US95-10398-34 - 30.00 107.07 192.62 576  
 seq\_name: /cgn2\_6/ptodata/1/ina/6/COMB.seq:US-08-422-108-2

seq\_documentation\_block:

; Sequence 2, Application US/08422108  
 ; Patent No. 6015567  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Hudziak, Robert M.  
 ; APPLICANT: Shepard, H. Michael  
 ; APPLICANT: Ullrich, Axel  
 ; TITLE OF INVENTION: HER2 EXTRACELLULAR DOMAIN  
 ; NUMBER OF SEQUENCES: 2  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Genentech, Inc.  
 ; STREET: 460 Point San Bruno Blvd  
 ; CITY: South San Francisco  
 ; STATE: California  
 ; COUNTRY: USA  
 ; ZIP: 94080  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: WinPatIn (Genentech)  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/422,108  
 ; FILING DATE: 14-Apr-1995  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 08/355460  
 ; FILING DATE: 13-DEC-1994  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 08/048346  
 ; FILING DATE: 15-APR-1993  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 07/354319  
 ; FILING DATE: 19-MAY-1989  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Lee, Wendy M  
 ; REGISTRATION NUMBER: 00,000  
 ; REFERENCE/DOCKET NUMBER: 554C2D2  
 ; TELEPHONE: 415/225-1994  
 ; TELEFAX: 415/952-9881  
 ; TELEX: 910/371-7168  
 ; INFORMATION FOR SEQ ID NO: 2:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1872 nucleotides  
 ; TYPE: Nucleic Acid  
 ; STRANDEDNESS: Single  
 ; TOPOLOGY: Linear  
 ; US-08-422-108-2

alignment\_scores:  
 Quality: 43.00 Length: 9  
 Ratio: 4.778 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-277-074-10 x US-08-422-108-2  
 Align seg 1/1 to: US-08-422-108-2 from: 1 to: 1872  
 1 LysilePheGlySerLeuAlaPheLeu 9  
 |||||  
 1042 AAGAUUUUGGAGCCUGCAUUCUG 1068

seq\_name: /cgn2\_6/ptodata/1/ina/5D\_COMB.seq:US-09-146-283-3

## seq\_documentation\_block:

; Sequence 3, Application US/09146283  
; Patent No. 5976546  
; GENERAL INFORMATION:  
; APPLICANT: Laus, Reiner  
; APPLICANT: Ruegg, Curtis L.  
; APPLICANT: Wu, Hongyu  
; TITLE OF INVENTION: Immunostimulatory Compositions  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dehlinger & Associates  
; STREET: 350 Cambridge Ave. Suite 250  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94306

## COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/146,283  
; FILING DATE: 03-SEPT-1998  
; CLASSIFICATION: 536  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Judge, Linda R.  
; REGISTRATION NUMBER: 42,702  
; REFERENCE/DOCKET NUMBER: 7636-0010.21  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 650-324-0880  
; TELEFAX: 650-324-0960  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2385 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: homo sapiens  
; INDIVIDUAL ISOLATE: GM-CSF-HER-2 fusion gene; Fig. 8  
; US-09-146-283-3

alignment\_scores:  
Quality: 43.00 Length: 9  
Ratio: 4.778 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-09-277-074-10 x US-09-146-283-3 ..  
Align seg 1/1 to: US-09-146-283-3 from: 1 to: 2385

## 1 LysilePheGlySerLeuAlaPheLeu 9

|||||  
1115 AGATCTTTGGGAGCGCTGCAATTCTG 1141

seq\_name: /cgn2\_6/ptodata/1/ina/6\_COMB.seq:US-08-579-823A-3

## seq\_documentation\_block:

; Sequence 3, Application US/08579823A  
; Patent No. 6080409  
; GENERAL INFORMATION:  
; APPLICANT: Laus, Reiner  
; APPLICANT: Ruegg, Curtis L.  
; APPLICANT: Wu, Hongyu  
; TITLE OF INVENTION: Immunostimulatory Composition and Method

; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dehlinger & Associates  
; STREET: 350 Cambridge Ave. Suite 250  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94306  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/579,823A  
; FILING DATE: 03-DEC-1998  
; CLASSIFICATION: 536  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Judge, Linda R.  
; REGISTRATION NUMBER: 42,702  
; REFERENCE/DOCKET NUMBER: 7636-0010  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 650-324-0880  
; TELEFAX: 650-324-0960  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2385 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: homo sapiens  
; INDIVIDUAL ISOLATE: GM-CSF-HER-2 fusion gene; Fig. 8  
; US-08-579-823A-3

## alignment\_scores:

Quality: 43.00 Length: 9  
Ratio: 4.778 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-09-277-074-10 x US-08-579-823A-3 ..  
Align seg 1/1 to: US-08-579-823A-3 from: 1 to: 2385

## 1 LysilePheGlySerLeuAlaPheLeu 9

|||||  
1115 AGATCTTTGGGAGCGCTGCAATTCTG 1141

seq\_name: /cgn2\_6/ptodata/1/ina/5C\_COMB.seq:US-08-625-101-1

## seq\_documentation\_block:

; Sequence 1, Application US/08625101  
; Patent No. 5869445  
; GENERAL INFORMATION:  
; APPLICANT: Cheever, Martin A.  
; APPLICANT: Disis, Mary L.

; TITLE OF INVENTION: COMPOUNDS FOR ELICITING OR ENHANCING IMMUNE  
; TITLE OF INVENTION: REACTIVITY TO HER-2/neu PROTEIN FOR PREVENTION  
; TITLE OF INVENTION: OR TREATMENT OF MALIGNANCIES IN WHICH THE HER-2/neu  
; TITLE OF INVENTION: ONCOGENE IS ASSOCIATED  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SEED and BERRY LLP  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98104-7092



```
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: 01-APR-1996
; APPLICATION NUMBER: US/08/625,101
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Sharkey, Richard G.
; REGISTRATION NUMBER: 32,629
; REFERENCE/DOCKET NUMBER: 920010.448C7
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3768 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..3765
; US-08-625-101-1

alignment_scores:
  Quality: 43.00      Length: 9
  Ratio: 4.778       Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-277-074-10 x US-08-625-101-1 ..
Align seg 1/1 to: US-08-625-101-1 from: 1 to: 3768

1 LysillepheglySerLeuAlaPheLeu 9
|||||
1105 AAGATCTTTGGGAGCGCTGGCATTTCTG 1131

seq_name: /cgn2_6/ptodata/1/lna/5C_COMB.seq:US-08-356-786-1

seq_documentation_block:
; Sequence 1, Application US/08356786
; Patent No. 5877305
; GENERAL INFORMATION:
; APPLICANT: Huston, James S.
; APPLICANT: Oppermann, Hermann
; APPLICANT: Houston, L. L.
; APPLICANT: King, David B.
; TITLE OF INVENTION: Biosynthetic Binding Protein for Cancer
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Edmund R. Pitcher, Testa, Hurwitz, & Thibault
; STREET: Exchange Place, 53 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/356,786
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/831,967
```

```
;
; FILING DATE: 06-FEB-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Pitcher, Edmund R.
; REGISTRATION NUMBER: 27,829
; REFERENCE/DOCKET NUMBER: CRP-053
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 248-7000
; TELEFAX: (617) 248-7100
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3768 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..3768
; OTHER INFORMATION: /note= "product = "cerB-b2""
; US-08-356-786-1

alignment_scores:
  Quality: 43.00      Length: 9
  Ratio: 4.778       Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-277-074-10 x US-08-356-786-1 ..
Align seg 1/1 to: US-08-356-786-1 from: 1 to: 3768

1 LysillepheglySerLeuAlaPheLeu 9
|||||
1105 AAGATCTTTGGGAGCGCTGGCATTTCTG 1131

seq_name: /cgn2_6/ptodata/1/lna/5A_COMB.seq:US-08-229-515A-14

seq_documentation_block:
; Sequence 14, Application US/08229515A
; Patent No. 5518885
; GENERAL INFORMATION:
; APPLICANT: RAZIUDIN
; APPLICANT: SARKAR, FAZLUL H
; TITLE OF INVENTION: ERBB2 PROMOTER BINDING PROTEIN IN
; TITLE OF INVENTION: NEOPLASTIC DISEASE
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NEEDLE & ROSENBERG PC
; STREET: 127 Peachtree Street, Suite 1200
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/229,515A
; FILING DATE: 19 APR 1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: PERRYMAN, DAVID G
; REGISTRATION NUMBER: 33,438
; REFERENCE/DOCKET NUMBER: 1414.608
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404-688-0770
; TELEFAX: 404-688-9880
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3955 base pairs
```

```
;
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-229-515A-14

alignment_scores:
  Quality: 43.00      Length: 9
  Ratio: 4.778       Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
US-09-277-074-10 x US-08-229-515A-14 ..
Align seg 1/1 to: US-08-229-515A-14 from: 1 to: 3955

1 LysilePheGlySerLeuAlaPheLeu 9
|||||
1133 AAGATCTTTGGGAGCTGGCATTTTG 1159

seq_name: /cgn2_6/ptodata/1/lna/5A_COMB.seq:US-08-645-865-14

seq_documentation_block:
; Sequence 14, Application US/08645865
; Patent No. 5654406
; GENERAL INFORMATION:
; APPLICANT: RAZIUDIN
; APPLICANT: SARKAR, FAZLUL H
; TITLE OF INVENTION: ERBB2 PROMOTER BINDING PROTEIN IN
; TITLE OF INVENTION: NEOPLASTIC DISEASE
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NEEDLE & ROSENBERG PC
; STREET: 127 Peachtree Street, Suite 1200
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: usa
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/645,865
; FILING DATE: 14 MAY 1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: PERRYMAN, DAVID G
; REGISTRATION NUMBER: 33,438
; REFERENCE/DOCKET NUMBER: 1414.608
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404-688-0770
; TELEFAX: 404-688-9880
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3955 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-645-865-14

alignment_scores:
  Quality: 43.00      Length: 9
  Ratio: 4.778       Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
US-09-277-074-10 x US-08-645-865-14 ..
Align seg 1/1 to: US-08-645-865-14 from: 1 to: 3955
```

```
1 LysilePheGlySerLeuAlaPheLeu 9
|||||
1133 AAGATCTTTGGGAGCTGGCATTTTG 1159

seq_name: /cgn2_6/ptodata/1/lna/5D_COMB.seq:US-09-048-804-1

seq_documentation_block:
; Sequence 1, Application US/09048804
; Patent No. 5968748
; GENERAL INFORMATION:
; APPLICANT: C. Frank Bennett, Allan Lipton, Lois M. Witters
; TITLE OF INVENTION: ANTISENSE OLIGONUCLEOTIDE MODULATION OF
; TITLE OF INVENTION: HUMAN HER-2 EXPRESSION
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5968748ris LLP
; STREET: One Liberty Place - 46th Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: U.S.A.
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 1.44 Mb diskette
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/048,804
; FILING DATE: Herewith
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Paul K. Legaard
; REGISTRATION NUMBER: 38,534
; REFERENCE/DOCKET NUMBER: ISIS-2913
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-3100
; TELEFAX: (215) 568-3439
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4473 base pairs
; TYPE: Nucleic Acid
; STRANDEDNESS: Single
; TOPOLOGY: Unknown
; ANTI-SENSE: No
US-09-048-804-1

alignment_scores:
  Quality: 43.00      Length: 9
  Ratio: 4.778       Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
US-09-277-074-10 x US-09-048-804-1 ..
Align seg 1/1 to: US-09-048-804-1 from: 1 to: 4473

1 LysilePheGlySerLeuAlaPheLeu 9
|||||
1279 AAGATCTTTGGGAGCTGGCATTTCTG 1305

seq_name: /cgn2_6/ptodata/1/lna/5A_COMB.seq:US-08-229-515A-9

seq_documentation_block:
; Sequence 9, Application US/08229515A
; Patent No. 5518885
; GENERAL INFORMATION:
; APPLICANT: RAZIUDIN
; APPLICANT: SARKAR, FAZLUL H
; TITLE OF INVENTION: ERBB2 PROMOTER BINDING PROTEIN IN
; TITLE OF INVENTION: NEOPLASTIC DISEASE
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
```

```
; ADDRESSEE: NEEDLE & ROSENBERG PC
; STREET: 127 Peachtree Street, Suite 1200
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: usa
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/229,515A
; FILING DATE: 19 APR 1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: PERRYMAN, DAVID G
; REGISTRATION NUMBER: 33,438
; REFERENCE/DOCKET NUMBER: 1414.608
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404-688-0770
; TELEFAX: 404-688-9880
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4530 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-229-515A-9

alignment_scores:
  Quality: 43.00      Length: 9
  Ratio: 4.778       Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
  US-09-277-074-10 x US-08-229-515A-9 ..
  Align seg 1/1 to: US-08-229-515A-9 from: 1 to: 4530
  1 LyslePheGlySerLeuAlaPheLeu 9
  |||||||||||||||||||||||||||
  1255 AAGATCTTTGGAGCTGGCATTCTG 1281

seq_name: /cgn2_6/ptodata/1/lna/5A_COMB.seq:US-08-645-865-9

seq_documentation_block:
; Sequence 9, Application US/08645865
; Patent No. 5654406
; GENERAL INFORMATION:
; APPLICANT: RAZIUDIN
; APPLICANT: SARKAR, FAZLUL H
; TITLE OF INVENTION: ERB2 PROMOTER BINDING PROTEIN IN
; TITLE OF INVENTION: NEOPLASTIC DISEASE
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NEEDLE & ROSENBERG PC
; STREET: 127 Peachtree Street, Suite 1200
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: usa
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/645,865
; FILING DATE: 14 MAY 1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
```

```
; NAME: PERRYMAN, DAVID G
; REGISTRATION NUMBER: 33,438
; REFERENCE/DOCKET NUMBER: 1414.608
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404-688-0770
; TELEFAX: 404-688-9880
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4530 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-645-865-9

alignment_scores:
  Quality: 43.00      Length: 9
  Ratio: 4.778       Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
  US-09-277-074-10 x US-08-645-865-9 ..
  Align seg 1/1 to: US-08-645-865-9 from: 1 to: 4530
  1 LyslePheGlySerLeuAlaPheLeu 9
  |||||||||||||||||||||||||||
  1255 AAGATCTTTGGAGCTGGCATTCTG 1281

seq_name: /cgn2_6/ptodata/1/lna/5B_COMB.seq:US-08-832-883-50

seq_documentation_block:
; Sequence 50, Application US/08832883
; Patent No. 5807681
; GENERAL INFORMATION:
; APPLICANT: Giordano, Antonio
; APPLICANT: Baldi, Alphonso
; TITLE OF INVENTION: METHODS FOR THE DIAGNOSIS AND PROGNOSIS
; TITLE OF INVENTION: OF CANCER
; NUMBER OF SEQUENCES: 115
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEIDEL, GONDA, LAVORGNA & MONACO, P.C.
; STREET: Suite 1800 Two Penn Center Plaza
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/832,883
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Monaco, Daniel A
; REGISTRATION NUMBER: 30,480
; REFERENCE/DOCKET NUMBER: 8321-13 US1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-8383
; TELEFAX: (215) 568-5549
; INFORMATION FOR SEQ ID NO: 50:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1618 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-832-883-50
```

```
alignment_scores:
  Quality: 33.00      Length: 9
  Ratio: 4.125       Gaps: 0
  Percent Similarity: 88.889  Percent Identity: 77.778

alignment_block:
US-09-277-074-10 x US-08-832-883-50  ..
Align seg 1/1 to: US-08-832-883-50 from: 1 to: 1618

1 LysillePheGlySerLeuAlaPheLeu 9
|||||
17 AGATTTTGGCAATCTCGTTCTG 43

seq_name: /cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-832-877-50

seq_documentation_block:
; Sequence 50, Application US/08832877
; Patent No. 5840506
; GENERAL INFORMATION:
; APPLICANT: Giordano, Antonio
; TITLE OF INVENTION: METHODS FOR THE DIAGNOSIS AND PROGNOSIS OF
; TITLE OF INVENTION: CANCER
; NUMBER OF SEQUENCES: 116
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEIDEL, GONDA, LAVORGNA & MONACO, P.C.
; STREET: Suite 1800 Two Penn Center Plaza
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/832,877
; FILING DATE:
; CLASSIFICATION: 436
; ATTORNEY/AGENT INFORMATION:
; NAME: Monaco, Daniel A
; REGISTRATION NUMBER: 30,480
; REFERENCE/DOCKET NUMBER: 8321-13 US2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-8383
; TELEFAX: (215) 568-5549
; INFORMATION FOR SEQ ID NO: 50:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1618 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-832-877-50

alignment_scores:
  Quality: 33.00      Length: 9
  Ratio: 4.125       Gaps: 0
  Percent Similarity: 88.889  Percent Identity: 77.778

alignment_block:
US-09-277-074-10 x US-08-832-877-50  ..
Align seg 1/1 to: US-08-832-877-50 from: 1 to: 1618

1 LysillePheGlySerLeuAlaPheLeu 9
|||||
17 AGATTTTGGCAATCTCGTTCTG 43

seq_name: /cgn2_6/ptodata/1/ina/6_COMB.seq:US-08-975-703-5
```

```
seq_documentation_block:
; Sequence 5, Application US/08975703
; Patent No. 6030832
; GENERAL INFORMATION:
; APPLICANT: Wong, Alexander K.C.
; APPLICANT: Bartel, Paul L.
; APPLICANT: Teng, David H.-F.
; APPLICANT: Tavtigian, Sean V.
; TITLE OF INVENTION: A Carboxy-Terminal BRCA1 Interacting
; TITLE OF INVENTION: Protein
; NUMBER OF SEQUENCES: 41
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Rothwell, Figg, Ernst & Kurz, P.C.
; STREET: 555 Thirteenth Street, N.W., Suite 701 East
; STREET: Tower
; CITY: Washington
; STATE: DC
; COUNTRY: U.S.A.
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/975,703
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Saxe, Stephen A.
; REGISTRATION NUMBER: 38,609
; REFERENCE/DOCKET NUMBER: 2318-0174
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-624-1589
; TELEFAX: 202-783-6031
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2694 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..2691
US-08-975-703-5

alignment_scores:
  Quality: 32.00      Length: 9
  Ratio: 3.556       Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 66.667

alignment_block:
US-09-277-074-10 x US-08-975-703-5  ..
Align seg 1/1 to: US-08-975-703-5 from: 1 to: 2694

1 LysillePheGlySerLeuAlaPheLeu 9
|||||
2546 AGAATTTTGGAAAGTTGGTTTCCTT 2572

seq_name: /cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-426-627-3

seq_documentation_block:
; Sequence 3, Application US/08426627
; Patent No. 5756664
; GENERAL INFORMATION:
; APPLICANT: Amann, Egon
```

APPLICANT: Otawara-Hamamoto, Yoko  
APPLICANT: Kikuno, Reiko  
APPLICANT: Takeshita, Sunao  
APPLICANT: Tezuka, Kenichi  
TITLE OF INVENTION: No. 575664el Protein with Bone Formation  
TITLE OF INVENTION: Ability and Process for its Production.  
NUMBER OF SEQUENCES: 24  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &  
ADDRESSEE: Dunner  
STREET: 1300 I Street, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20005-3315  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/426,627  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/036,841  
FILING DATE: 25-MAR-1993  
APPLICATION NUMBER: JP 4-71501  
FILING DATE: 27-MAR-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Hammond, Alan W.  
REGISTRATION NUMBER: 35,178  
REFERENCE/DOCKET NUMBER: 02481-1285-00000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-408-4000  
TELEFAX: 202-408-4400  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3092 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA to mRNA  
ORGANISM: Homo sapiens  
TISSUE TYPE: Placenta  
FEATURE:  
NAME/KEY: CDS  
LOCATION: join(38..2375)  
FEATURE:  
NAME/KEY: mat\_peptide  
LOCATION: join(101..2375)  
US-08-426-627-3

alignment\_scores:  
Quality: 32.00 Length: 9  
Ratio: 3.556 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 66.667

alignment\_block:  
US-09-277-074-10 x US-08-426-627-3/rev ..

Align seg 1/1 to reverse of: US-08-426-627-3 from: 1 to: 3092

1 LysilePheGlySerLeuAlaPheLeu 9  
|||||:|||||:|||||:|||||  
323 AAACGTGGGCGACGCTTCATTCCTT 297

seq\_name: /cgn2\_6/ptodata/1/lna/5C\_COMB.seq:US-08-477-396A-3

seq\_documentation\_block:

; Sequence 3, Application US/08477396A

Patent No. 5872235  
GENERAL INFORMATION:  
APPLICANT: Chen, Lan Bo  
APPLICANT: Bao, Shideng  
APPLICANT: Liu, Yuan  
TITLE OF INVENTION: A NOVEL TUMOR MARKER AND NOVEL METHOD OF  
TITLE OF INVENTION: ISOLATING SAME  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Weingarten, Schurgin, Gagnebin & Hayes  
STREET: Ten Post Office Square  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/477,396A  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/146,488  
FILING DATE: 29-OCT-1993  
APPLICATION NUMBER: US 08/448,388  
FILING DATE: 28-MAY-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US94/12502  
FILING DATE: 31-OCT-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Heine, Holliday C.  
REGISTRATION NUMBER: 34,346  
REFERENCE/DOCKET NUMBER: DFCI-3338X  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 542-2290  
TELEFAX: (617) 451-0313  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3126 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 43..2376  
US-08-477-396A-3

alignment\_scores:  
Quality: 32.00 Length: 9  
Ratio: 3.556 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 66.667

alignment\_block:  
US-09-277-074-10 x US-08-477-396A-3/rev ..

Align seg 1/1 to reverse of: US-08-477-396A-3 from: 1 to: 3126

1 LysilePheGlySerLeuAlaPheLeu 9  
|||||:|||||:|||||:|||||  
346 AAACGTGGGCGACGCTTCATTCCTT 320



OM of: US-09-277-074-10 to: EST:\* out\_format : pfs

Date: Nov 15, 2000 9:56 PM

About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=framed\_p2n.model -DEV=xlp  
-O=/cgn1\_1/USPfo\_Spool/US09277074/runat\_14112000\_120308\_22605/app\_query.fasta\_1.67  
-DB=EST -OPT=fastap -SUFFIX=rst -GAPOP=12.000 -GAPEXT=4.000  
-MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000 -GAPOP=4.500  
-GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -GAPOP=6.000  
-GAPEXT=7.000 -XGAPOP=10.000 -XGAPEXT=0.500 -DELOP=6.000  
-DELEXT=7.000 -START=1 -MATRIX=blotsum62 -TRANS=human40.cdi  
-LIST=45 -LOCAL=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0  
-ALIGN=15 -MODE=LOCAL -OUTFT=pfs -NORM=ext -MINLEN=0  
-MAXLEN=200000000 -USER=US09277074\_ECGN1\_1\_1780 -NCPU=6  
-ICPU=3 -LONGLOG -NO\_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-277-074-10

Query length: 9

Database: EST\*

Database sequences: 7189864

Database length: -1203564053

Search time (sec): 841.950000

score\_list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
gb_est15:AV047157	+	39.00	144.11	46.23	AV047157 AV047157 Mus musculus
gb_est20:AW228360	-	38.00	146.10	35.82	AW228360 up20h09.yi NCI_CGAP_Ma
gb_est22:AA213247	-	38.00	140.92	69.57	AA213247 mw80f09.x1 Soares mus
gb_est19:AW047802	-	38.00	137.51	107.87	AW047802 UI-M-BH1-alk-g-12-0-UI
gb_est16:AA792385	-	38.00	137.33	110.37	AA792385 vp8f01.r1 Stratagene
gb_est33:BE137507	-	38.00	137.17	112.57	BE137507 ug64c05.y1 Soares_mamm
gb_est35:BE448095	-	38.00	137.17	112.57	BE448095 ut82h07.y1 Soares_mamm
gb_est1:AA097521	-	38.00	135.51	139.41	AA097521 mk15f06.r1 Soares_mamm
gb_est35:AA646381	-	38.00	135.08	147.24	AA646381 vni4g01.r1 Stratagene
gb_est33:BE134483	-	38.00	134.46	159.46	BE134483 ugi1e08.y1 Soares_mamm
gb_gss24:CN5055MV	-	38.00	128.91	324.81	AL322240 Tetradodon nigroviride
gb_est31:BB472014	-	37.00	139.67	81.68	BB472014 BB472014 RIKEN full-le
gb_est15:AV029366	-	37.00	139.38	84.87	AV029366 AV029366 Mus musculus
gb_est23:AW349836	+	37.00	137.45	108.62	AW349836 L0055H09.3 Mouse E12.5
gb_est25:AW906929	+	37.00	136.64	120.60	AW906929 EST343052 potato stolo
gb_gss14:AZ009927	-	37.00	135.01	148.62	AW009927 RPCI-23-324N2.TJ RPCI-
gb_est24:AW837616	-	37.00	134.98	149.12	AW837616 QV2-LT0039-250300-100-
gb_est10:AI382909	-	37.00	133.97	169.84	AI382909 tc20d02.x1 Soares_Nhm
gb_gss1:AO001963	-	37.00	132.20	213.03	AO001963 CIT-HSP-2283P20.TR CIT
gb_est35:BE339777	+	37.00	131.69	227.41	BE339777 EST343836 potato stolo
gb_est24:AW837546	+	37.00	131.45	234.39	AW837546 QV2-LT0039-250300-099-
gb_est24:AW837513	+	37.00	130.98	248.98	AW837513 QV2-LT0038-270300-108-
gb_est24:AW837516	+	37.00	130.33	270.86	AW837516 QV2-LT0038-270300-108-
gb_est24:AW837506	+	37.00	129.98	283.02	AW837506 EST362761 MAGE resequ
gb_est24:AW837506	+	37.00	129.29	309.29	AW837506 QV2-LT0038-270300-108-
gb_est26:BB054043	+	36.00	134.86	151.43	BB054043 BB054043 RIKEN full-le
gb_est30:BB379144	+	36.00	134.22	164.49	BB379144 BB379144 RIKEN full-le
gb_est37:H74306	-	36.00	134.00	169.14	H74306 yu56d09.s1 Soares fetal
gb_est15:AU058062	-	36.00	133.93	170.69	AU058062 AU058062 Rice panicle
gb_est17:AV242040	+	36.00	133.89	171.47	AV242040 AV242040 RIKEN full-le
gb_est14:AI944779	+	36.00	131.75	225.62	AI944779 bs05d06.y1 Drosophila
gb_est39:RI5799	+	36.00	130.64	260.20	RI5799 IB1906 Infant brain, Ben
gb_est16:AV191538	+	36.00	130.52	264.37	AV191538 AV191538 Yuji Kohara u
gb_est5:AA701356	+	36.00	130.30	271.90	AA701356 z163e12.s1 Soares_feta
gb_gss7:AQ496013	+	36.00	129.85	287.91	AQ496013 HS_5064.B1.D06.SP6E.RP
gb_est39:T66861	-	36.00	129.56	298.94	T66861 ya50d10.s2 Soares fetal
gb_est33:AA367800	-	36.00	129.43	304.06	AA367800 EST78936 Placenta 1 Hc
gb_est6:AA779245	+	36.00	129.25	310.90	AA779245 zj39g04.s1 Soares_feta
gb_est14:AU020415	+	36.00	129.21	312.62	AU020415 AU020415 Mouse eight-d
gb_gss13:AQ912881	+	36.00	128.89	325.53	AQ912881 nbab0038N12f CUGI Rice
gb_est11:AI510528	-	36.00	128.79	329.85	AI510528 mq11g06.y1 Barstead ME

gb\_est38:NG3803 - 36.00 128.33 349.84 478 ! N63803 za25g09.s1 Soares fet  
gb\_gss2:AQ134689 + 36.00 128.23 354.22 483 ! AQ134689 HS\_3052.B1.A09.MF C  
gb\_gss10:AQ171003 + 36.00 128.18 356.84 486 ! AQ171003 HS\_2152.B2.D08.MR C  
gb\_est20:AW230326 + 36.00 128.14 358.60 488 ! AW230326 uo62c03.y1 NCI\_CGAP

seq\_name: gb\_est15:AV047157

seq\_documentation\_block: 309 bp mRNA EST 23-NOV-1999  
LOCUS AV047157 Mus musculus adult C57BL/6J testis Mus musculus CDNA clone  
DEFINITION AV047157 Mus musculus adult C57BL/6J testis Mus musculus CDNA clone  
1700065P12, mRNA sequence.

ACCESSION AV047157

VERSION AV047157.2 GI:4866822

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus;

1 (bases 1 to 309)

REFERENCE

AUTHORS

Carinci, P., Shibata, K., Ozawa, Y., Konno, H., Itoh, M., Aizawa, K.,

Akahira, S., Akiyama, J., Fukuda, S., Fukunishi, Y., Funayama, T., Hara

, A., Hayatsu, N., Hori, F., Ishikawa, T., Itoh, M., Izawa, M., Kawai, J.,

Kikuchi, N., Kojima, Y., Matsuyama, T., Niitsuma, H., Oda, H., Owa, C.,

Sato, K., Shibata, Y., Shigemoto, Y., Shiraki, T., Sogabe, Y., Sugahara

, Y., Suzuki, H., Suzuki, H., Tateno, M., Tomaru, Y., Tomioka, N.,

Watanabe, S., Yagame, M., Yamamura, T., Yokota, T., Yoshino, M.,

Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.

RIKEN Mouse ESTs

Unpublished (1999)

Contact: Chie Owa

Genome Science Laboratory

RIKEN

3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan

Tel: 81-298-36-9145

Fax: 81-298-36-9098

Email: genome-resetc.riken.go.jp

Thermotabilization and thermoactivation of thermostable enzymes by

trehalose and its application for the synthesis of full length cDNA

(Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))

Transcriptional sequencing: A method for DNA sequencing using RNA

polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))

Please visit our web site (http://genome.rtc.riken.go.jp) for

further details.

FEATURES

Location/Qualifiers

1..309

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="1700065P12"

/sex="male"

/tissue\_type="testis"

/dev\_stage="adult"

BASE COUNT 70 a 50 c 70 g 119 t

ORIGIN

alignment\_scores:

Quality: 39.00 Length: 9

Ratio: 4.333 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 77.778

alignment\_block:

US-09-277-074-10 x AV047157

Align seg 1/1 to: AV047157 from: 1 to: 309

1 LysillephecysSerleualapheteu 9

220 AGGGTGTGGTTCCTGGCATTTTG 246

seq\_name: gb\_est20:AW228360

```

seq_documentation_block: 164 bp mRNA EST 10-DEC-1999
LOCUS AW228360 Mus musculus cDNA clone IMAGE:2654945 5',
DEFINITION up20h09.y1 NCI_CGAP_Mam2 Mus musculus cDNA clone IMAGE:2654945 5',
mRNA sequence.
ACCESSION AW228360
VERSION AW228360.1 GI:6557656
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 164)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
MGI:1033093
Seq primer: -40RP from Gibco
High quality sequence stop: 157.
FEATURES
source
Location/Qualifiers
1..164
/organism="Mus musculus"
/strain="FVB-3"
/db_xref="taxon:10090"
/clone="IMAGE:2654945"
/clone_lib="NCI_CGAP_Mam2"
/tissue_type="tumor, biopsy sample"
/dev_stage="5 months"
/lab_host="DH10B"
/notes="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Library constructed by Life Technologies. Investigator
providing samples: Gilbert Smith, NIH"
BASE COUNT 40 a 55 c 30 g 39 t
ORIGIN
alignment_scores:
Quality: 38.00 Length: 8
Ratio: 4.750 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 87.500
alignment_block:
US-09-277-074-10 x AW228360/rev ..
Align seg 1/1 to reverse of: AW228360 from: 1 to: 164
seq_name: gb_est2:AA213247
1 LysilePheGlySerLeuAlaPhe 8
||||:|||||
158 AAGGTGTTTGGAGTTTGGCCTTC 135
seq_name: gb_est2:AA213247
seq_documentation_block: 286 bp mRNA EST 18-FEB-1997
LOCUS AA213247 Soares mouse NML Mus musculus cDNA clone IMAGE:677033
DEFINITION mw80f09.r1 Soares mouse NML Mus musculus cDNA clone IMAGE:677033
5', mRNA sequence.
ACCESSION AA213247
VERSION AA213247.1 GI:1811918
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 413)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene

```

```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 286)
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
The WashU-HMI Mouse EST Project
Unpublished (1996)
Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:416737
Seq primer: -28mi3 rev2 ET from Amersham
High quality sequence stop: 281.
FEATURES
source
Location/Qualifiers
1..286
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone_lib="Soares mouse NML"
/tissue_type="Liver"
/lab_host="DH10B"
/notes="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dt) primer [5'
TGTTCACCAATCTGAAGTCGAGCGCCGCAATCTTTTTTTTTTTT 3'];
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
constructed and normalized by Bento Soares and M.Fatima
Bonaldo."
BASE COUNT 67 a 101 c 43 g 75 t
ORIGIN
alignment_scores:
Quality: 38.00 Length: 8
Ratio: 4.750 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 87.500
alignment_block:
US-09-277-074-10 x AA213247/rev ..
Align seg 1/1 to reverse of: AA213247 from: 1 to: 286
seq_name: gb_est19:AW047802
1 LysilePheGlySerLeuAlaPhe 8
||||:|||||
248 AAGGTGTTTGGAGTTTGGCCTTC 225
seq_name: gb_est19:AW047802
seq_documentation_block: 413 bp mRNA EST 18-SEP-1999
LOCUS AW047802 NIH_BMAP_M_S2 Mus musculus cDNA clone
DEFINITION UI-M-BH1-alk-g-12-0-UI.s1 NIH_BMAP_M_S2 Mus musculus cDNA clone
UI-M-BH1-alk-g-12-0-UI 3', mRNA sequence.
ACCESSION AW047802
VERSION AW047802.1 GI:5908331
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 413)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene

```



JOURNAL  
MEDLINE  
COMMENT

discovery  
Genome Res. 6 (9), 791-806 (1996)  
97044477  
Contact: Chin, H  
National Institute of Mental Health  
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD  
20892-9643, USA  
Tel: 301 443 1706  
Fax: 301 443 9890  
Email: mES@mail.nih.gov  
The sequence contained an oligo-dT track that was present in the  
oligonucleotide that was used to prime the synthesis of first  
strand cDNA and therefore this may represent a bonafide poly A  
tail. The sequence tag present in the cDNA between the NotI site  
and the oligo-dT track served to identify it as a clone from the  
normalized corpus striatum library cDNA Library Preparation: M.B.  
Soares Lab Clone Distribution: NIH BMAP cDNA clones will be made  
available by the means that is soon to be determined. When NIH  
determines the means for distribution of the BMAP cDNA clones, this  
record will be updated accordingly when that means is determined.  
The following repetitive elements were found in this cDNA sequence:  
1-64, >BCL1\_MM4scrNA  
Seq primer: M13 Forward  
POLYA=Yes.

FEATURES  
source  
1. .413  
Location/Qualifiers  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UI-M-BH1-alk-g-12-0-UI"  
/clone\_lib="NIH\_BMAP\_M\_S2"  
/dev\_stage="27-32 days"  
/lab\_host="DH10B (Life Technologies)"  
/note="vector: pT73D-Pac (Pharmacia) with a modified  
polylinker; Site\_1: Not I; Site\_2: Eco RI; The  
NIH\_BMAP\_M\_S2 library is a subtracted library derived from  
NIH\_BMAP\_M\_S1, which in turn is a subtracted library  
derived from a mixture of normalized libraries from ten  
regions of the mouse brain (cerebellum, brain stems,  
olfactory bulbs, hypothalamus, cortex, amygdala, basal  
ganglia, pineal gland, striatum, hippocampus). The driver  
used for subtraction consisted of a pool of 5,000 clones  
from the NIH\_BMAP\_M\_S1 library and a pool of 2,000 clones  
obtained from non-normalized and normalized mouse brain  
spinal cord libraries.  
TAG\_LIB="NIH\_BMAP\_M\_S2"  
TAG\_TISSUE="corpus-striatum"  
TAG\_SEQ="ACGGC"

BASE COUNT 78 a 71 c 119 g 144 t 1 others  
ORIGIN

alignment\_scores:  
Quality: 38.00 Length: 8  
Ratio: 4.750 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 87.500

alignment\_block:

US-09-277-074-10 x AW047802

Align seg 1/1 to: AW047802 from: 1 to: 413

1 LyslePheGlySerLeuAlaPhe 8  
|||||  
282 AAGGTGTTTGAAGTTGGCCTC 305

seq\_name: gb\_est6:AA792385

seq\_documentation\_block:

LOCUS AA792385 421 bp mRNA EST 09-FEB-1998  
DEFINITION vp88f01.r1 Stratagene mouse diaphragm (#937303) Mus musculus cDNA  
clone IMAGE:1091833 5', mRNA sequence.  
ACCESSION AA792385

VERSION AA792385.1 GI:2855340  
KEYWORDS EST  
SOURCE house mouse.  
ORGANISM Mus musculus

REFERENCE  
AUTHORS

1 (bases 1 to 421)  
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,  
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,  
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,  
Theising, B., Wyllie, T., Lennon, G., Soares, B., Wilson, R. and  
Waterston, R.

TITLE  
JOURNAL  
COMMENT

The WashU-HHMI Mouse EST Project  
Unpublished (1996)  
Contact: Marra M/Mouse EST Project  
WashU-HHMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810

Email: mouseest@watson.wustl.edu  
This clone is available royalty-free through LLNL: contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:598065

Seq primer: -28ml3 rev1 ET from Amersham  
High quality sequence stop: 390.

FEATURES  
source  
1. .421  
Location/Qualifiers

/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="IMAGE:1091833"  
/clone\_lib="Stratagene mouse diaphragm (#937303)"  
/tissue\_type="diaphragm"  
/dev\_stage="adult"  
/lab\_host="SOLR (kanamycin resistant)"  
/note="Organ: diaphragm; Vector: pBluescript SK-; Site\_1:  
EcoRI; Site\_2: XhoI; Cloned unidirectionally from mRNA  
prepared from diaphragm muscle. Primer: Oligo dT. Average  
insert size: 1.5 kb. Uni-ZAP XR Vector: -5' adaptor  
sequence: 5' GAATTCGGCAGCAG 3' -3' adaptor sequence: 5'  
CTCGAGTTTTTTTTTTTTTTT 3"

BASE COUNT 112 a 141 c 78 g 90 t  
ORIGIN

alignment\_scores:  
Quality: 38.00 Length: 8  
Ratio: 4.750 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 87.500

alignment\_block:

US-09-277-074-10 x AA792385/rev

Align seg 1/1 to reverse of: AA792385 from: 1 to: 421

1 LyslePheGlySerLeuAlaPhe 8  
|||||  
189 AAGGTGTTTGAAGTTGGCCTC 166

seq\_name: gb\_est33:BE137507

seq\_documentation\_block:

LOCUS BE137507 428 bp mRNA EST 21-JUN-2000  
DEFINITION ug64c05.y1 Soares\_mammmary\_gland\_NMLMG Mus musculus cDNA clone  
IMAGE:1547144 5', mRNA sequence.

ACCESSION BE137507

VERSION BE137507.1 GI:8600007

KEYWORDS EST.

SOURCE house mouse.

ORGANISM

Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 428)

**AUTHORS**  
**TITLE** NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
**JOURNAL**  
**COMMENT** Unpublished (1997)  
 Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 This clone is available royalty-free through LLNL ; contact the  
 IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
 MGI:952492  
 Seq primer: -40RP from Gibco  
 High quality sequence stop: 419.

**FEATURES**  
 source  
 Location/Qualifiers  
 1..428  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /clone\_lib="Soares\_mammary\_gland\_NMLMG"  
 /sex="female (lactating)"  
 /tissue\_type="mammary gland"  
 /lab\_host="DH10B"  
 /note="vector: p7T3D-Pac (Pharmacia) with a modified  
 polylinker; 1st strand cDNA was prepared from mammary  
 gland tissue from a lactating female, and was then primed  
 with a Not I - oligo(dT) primer. Double-stranded cDNA was  
 ligated to Eco RI adaptors (Pharmacia), digested with Not  
 I and cloned into the Not I and Eco RI sites of the  
 modified p7T3 vector. Library is normalized. Library  
 was constructed by Bento Soares and M. Fatima Bonaldo."  
 BASE COUNT 115 a 145 c 74 g 94 t  
 ORIGIN

alignment\_scores:  
 Quality: 38.00 Length: 8  
 Ratio: 4.750 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 87.500

alignment\_block:  
 US-09-277-074-10 x BE137507/rev ..  
 Align seg 1/1 to reverse of: BE137507 from: 1 to: 428

```

1 LysilePheGlySerLeuAlaphe 8
||||:|||||
202 AAGGTGTTGGAGTTGGCCTTC 179

```

seq\_name: gb\_est35:BE448085

seq\_documentation\_block: 428 bp mRNA EST 25-JUL-2000  
 LOCUS BE448085  
 DEFINITION ut82h07.y1 Soares\_mammary\_gland\_NMLMG Mus musculus cDNA clone  
 IMAGE:3368989 5', mRNA sequence.

ACCESSION BE448085  
 VERSION BE448085.1 GI:9447662  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus

REFERENCE  
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 428)

TITLE NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 JOURNAL National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 COMMENT Unpublished (1997)  
 Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 This clone is available royalty-free through LLNL ; contact the  
 IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
 MGI:1078593  
 Seq primer: -40RP from Gibco  
 High quality sequence stop: 427.

**FEATURES**  
 source  
 Location/Qualifiers  
 1..428  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /clone\_lib="Soares\_mammary\_gland\_NMLMG"  
 /sex="female (lactating)"  
 /tissue\_type="mammary gland"  
 /lab\_host="DH10B"  
 /note="vector: p7T3D-Pac (Pharmacia) with a modified  
 polylinker; 1st strand cDNA was prepared from mammary  
 gland tissue from a lactating female, and was then primed  
 with a Not I - oligo(dT) primer. Double-stranded cDNA was  
 ligated to Eco RI adaptors (Pharmacia), digested with Not  
 I and cloned into the Not I and Eco RI sites of the  
 modified p7T3 vector. Library is normalized. Library  
 was constructed by Bento Soares and M. Fatima Bonaldo."  
 BASE COUNT 111 a 150 c 76 g 91 t  
 ORIGIN

alignment\_scores:  
 Quality: 38.00 Length: 8  
 Ratio: 4.750 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 87.500

alignment\_block:  
 US-09-277-074-10 x BE448085/rev ..  
 Align seg 1/1 to reverse of: BE448085 from: 1 to: 428

```

1 LysilePheGlySerLeuAlaphe 8
||||:|||||
172 AAGGTGTTGGAGTTGGCCTTC 149

```

seq\_name: gb\_est1:AA097521

seq\_documentation\_block: 512 bp mRNA EST 25-OCT-1996  
 LOCUS AA097521  
 DEFINITION mk15f06.r1 Soares mouse p3NMF19.5 Mus musculus cDNA clone  
 IMAGE:492995 5', mRNA sequence.

ACCESSION AA097521  
 VERSION AA097521.1 GI:1643221  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus

REFERENCE  
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 512)

TITLE The WashU-HHMI Mouse EST Project  
 JOURNAL Unpublished (1996)  
 COMMENT Contact: Marra M/Mouse EST Project  
 WashU-HHMI Mouse EST Project  
 Washington University School  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@watson.wustl.edu  
 This clone is available royalty-free through LLNL ; contact the  
 IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
 MGI:296443  
 Seq primer: -28M13 rev2 from Amersham  
 High quality sequence stop: 457.

**FEATURES**  
 source  
 Location/Qualifiers  
 1..512  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:492995"

```

/clone_lib="Soares mouse p3NMF19.5"
/dev_stage="19.5 qpc total fetus"
/lab_host="PH10B (ampicillin resistant)"
/notes="Vector: pT73D (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5'
TCCTACCATCTGAAGTGGAGCGCGCATTTTTTTTTTT 3'].
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M. Fatima Bonaldo. RNA was kindly provided by
Dr. Minoru Ko (Wayne State University)."
BASE COUNT      127 a  173 c  105 g  106 t
ORIGIN

alignment_scores:
  Quality:      38.00      Length:      8
  Ratio:        4.750      Gaps:      0
Percent Similarity: 100.000 Percent Identity: 87.500

alignment_block:
US-09-277-074-10 x AA097521/rev ..
Align seg 1/1 to reverse of: AA097521 from: 1 to: 512

1 LysilePheGlySerLeuAlaPhe 8
||||:|||||
34 AAGGTGTTGGAAGTTGGCCTTC 11

seq_name: gb_est5:AA646381

seq_documentation_block:
LOCUS      AA646381      536 bp      mRNA      EST      28-OCT-1997
DEFINITION vn14g01.r1 Stratagene mouse heart (#937316) Mus musculus cDNA clone
IMAGE:1021200 5', mRNA sequence.
ACCESSION  AA646381
VERSION     AA646381.1 GI:2572810
KEYWORDS   EST.
SOURCE      house mouse.
ORGANISM    Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 536)
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
The WashU-HMI Mouse EST Project
Unpublished (1996)
Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:571976
Seq primer: -28m13 rev1 ET from Amersham
High quality sequence stop: 446.

FEATURES
  source
    1..536
    /organism="Mus musculus"
    /strain="NIH/Swiss"
    /db_xref="taxon:10090"
    /clone="IMAGE:1021200"
    /clone_lib="Stratagene mouse heart (#937316)"
    /sex="pooled"

```

```

/tissue_type="heart"
/dev_stage="13 day embryos"
/lab_host="SOUR (kanamycin resistant)"
/notes="Organ: heart; Vector: pluescript SK-; Site_1:
EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer:
Oligo dT. 93 pooled NIH/Swiss 13 day embryo hearts.
Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5'
adaptor sequence: 5' GAATTCGGCAGGAG 3' -3' adaptor
sequence: 5' CTCGAGTTTTTTTTTTTTTT 3'"
BASE COUNT      130 a  188 c  101 g  117 t
ORIGIN

alignment_scores:
  Quality:      38.00      Length:      8
  Ratio:        4.750      Gaps:      0
Percent Similarity: 100.000 Percent Identity: 87.500

alignment_block:
US-09-277-074-10 x AA646381/rev ..
Align seg 1/1 to reverse of: AA646381 from: 1 to: 536

1 LysilePheGlySerLeuAlaPhe 8
||||:|||||
66 AAGGTGTTGGAAGTTGGCCTTC 43

seq_name: gb_est33:BE134483

seq_documentation_block:
LOCUS      BE134483      573 bp      mRNA      EST      21-JUN-2000
DEFINITION ug19e08.y1 Soares_mammary_gland_NMLMG Mus musculus cDNA clone
IMAGE:1531718 5', mRNA sequence.
ACCESSION  BE134483
VERSION     BE134483.1 GI:8596983
KEYWORDS   EST.
SOURCE      house mouse.
ORGANISM    Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 573)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:1947818
Seq primer: -40RP from Gibco
High quality sequence stop: 462.

FEATURES
  source
    1..573
    /organism="Mus musculus"
    /db_xref="taxon:10090"
    /clone="IMAGE:1531718"
    /clone_lib="Soares_mammary_gland_NMLMG"
    /sex="female (lactating)"
    /tissue_type="mammary gland"
    /lab_host="DH10B"
    /notes="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; 1st strand cDNA was prepared from mammary
gland tissue from a lactating female, and was then primed
with a Not I - oligo(dT) primer. Double-stranded cDNA was
ligated to Eco RI adaptors (Pharmacia), digested with Not
I and cloned into the Not I and Eco RI sites of the
modified pT73 vector. Library is normalized. Library
was constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT      143 a  198 c  102 g  130 t
ORIGIN

```

alignment\_scores:  
 Quality: 38.00 Length: 8  
 Ratio: 4.750 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 87.500

alignment\_block:  
 US-09-277-074-10 x BE134483/rev ..

Align seg 1/1 to reverse of: BE134483 from: 1 to: 573

1 LyslephgGlySerLeuAlaPheLeu 8  
 |||:|||||:|||||:|||||:|||||

386 AGGTTGGAGTTGGCCTTC 363

seq\_name: gb\_gss24:CNS055MW

seq\_documentation\_block:  
 LOCUS CNS055MW 1040 bp DNA GSS 26-JUL-2000  
 DEFINITION Tetraodon nigroviridis genome survey sequence SP6 end of clone 001M14 of library B from Tetraodon nigroviridis, genomic survey sequence.  
 ACCESSION AL322240.1 GI:9555124  
 VERSION AL322240.1  
 KEYWORDS GSS: genome survey sequence.  
 SOURCE Tetraodon nigroviridis.  
 ORGANISM Tetraodon nigroviridis  
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Acanthopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodon.  
 REFERENCE 1 (bases 1 to 1040)  
 AUTHORS Roest Crolius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C., Fizames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J.  
 TITLE Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis  
 JOURNAL Genome Res. 10, 939-949 (2000)  
 REFERENCE 2 (bases 1 to 1040)  
 AUTHORS Roest Crolius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fischer,C., Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F., Saurin,W. and Weissenbach,J.  
 TITLE Estimate of human gene number provided by genome-wide analysis using Tetraodon nigroviridis DNA sequence  
 JOURNAL Nat. Genet. 25 (2), 235-238 (2000)  
 REFERENCE 3 (bases 1 to 1040)  
 AUTHORS Genoscope.  
 TITLE Direct Submission  
 JOURNAL Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases  
 COMMENT This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at <http://www.genoscope.cns.fr/tetraodon>.

FEATURES  
 source  
 Location/Qualifiers  
 1..1040  
 /organism="Tetraodon nigroviridis"  
 /db\_xref="taxon:99883"  
 /clone="001M14"  
 /clone\_lib="B"  
 /note="Genoscope sequence ID : COAB001BG07B1-end : SP6"

BASE COUNT 200 a 248 c 297 g 268 t 27 others  
 ORIGIN

alignment\_scores:  
 Quality: 38.00 Length: 8  
 Ratio: 4.750 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
 US-09-277-074-10 x CNS055MW

Align seg 1/1 to: CNS055MW from: 1 to: 1040

2 llepHeGlySerLeuAlaPheLeu 9  
 |||:|||||:|||||:|||||:|||||

344 ATCTTCGGGTCCTCGCTTCCTC 367

seq\_name: gb\_est31:BB472014

seq\_documentation\_block:  
 LOCUS BB472014 215 bp mRNA EST 22-JUL-2000  
 DEFINITION BB472014 RIKEN full-length enriched, 12 days embryo eyeball Mus musculus cDNA clone D230041P19 3', mRNA sequence.  
 ACCESSION BB472014  
 VERSION BB472014.1 GI:9389203  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 215)  
 AUTHORS Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N., Kiyosawa,H., Kojima,Y., Kondo,S., Koye,S., Kurihara,C., Kusakabe,M., Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y., Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata,Y., Shigenomi,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y., Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tominaga,N., Toya,T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,I., Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.  
 RIKEN Mouse ESTs (Konno,H., et al.)  
 UNPUBLISHED (2000)  
 CONTACT: Yoshihide Hayashizaki  
 Genome Exploration Research Group, Life Science Tsukuba Center,  
 Genome Science Laboratory  
 The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center  
 3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan  
 Tel: +81-298-36-9013  
 Fax: +81-298-36-9098  
 Email: genome.res@rtc.riken.go.jp  
 URL: <http://genome.rtc.riken.go.jp/>  
 Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.  
 Thermostabilization and thermoactivation of thermostable enzymes by trehalose and its application for the synthesis of full length cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)  
 Itoh,M., Kitsuunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J., Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.  
 Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)  
 Carninci,P. and Hayashizaki,Y.  
 High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)

Please visit our web site (<http://genome.rtc.riken.go.jp>) for further details.

FEATURES  
 source

Location/Qualifiers  
 1..215  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /clone="D230041P19"  
 /clone\_lib="RIKEN full-length enriched, 12 days embryo eyeball"  
 /tissue\_type="eyeball"  
 /dev\_stage="12 days embryo"  
 /lab\_host="DH10B"  
 /note="Site\_1: SalI; Site\_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in











OM of: US-09-277-074-10 to: GenEmbl.\* out\_format : pfs  
 Date: Nov 15, 2000 11:01 PM  
 About: Results were produced by the GenCore software, version 4.5.  
 Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:  
 -MODELframet\_p2n.model -DEV-xlp  
 -Q/cgcn2\_1/USPTO.spool/US09277074/runat\_14112000\_120415\_23077/app\_query.fasta\_1.67  
 -DB-GenEmbl -OFMT-fastap -SUFFIX=lim60.rge -GAPOP=12.000  
 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000  
 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500  
 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500  
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blom62  
 -TRANS-human40.cdi -LIST=45 -DCCALIGN=200 -THR\_SCORE=pct  
 -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTMT=pfs  
 -NORM-ext -MINLEN=0 -MAXLEN=60 -USER=US09277074.@CGCN1\_1.3727  
 -NCPU=6 -ICPU=3 -LONGLOG -NO\_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-277-074-10

Query length: 9

Database: GenEmbl.\*

Database sequences: 1033670

Database length: 2111177393

Search time (sec): 1109.630000

score\_list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
gb_pat: I03346	- 29.00	122.03	112.73	36	I03346 Sequence 4 from Patent US
gb_pat: A02190	- 28.00	119.58	154.53	30	A02190 Sequence 11 from Patent W
gb_pat: I20887	+ 28.00	116.05	242.80	47	I20887 Sequence 10 from Patent W
gb_pat: AR032803	- 28.00	115.57	258.40	50	AR032803 Sequence 415 from patent
gb_pat: I29543	- 28.00	115.57	258.40	50	I29543 Sequence 415 from patent
gb_pat: I91217	- 28.00	115.57	258.40	50	I91217 Sequence 415 from patent
gb_pat: AR014596	- 28.00	114.82	284.41	55	AR014596 Sequence 20 from patent
gb_pat: AR014597	+ 28.00	114.82	284.41	55	AR014597 Sequence 21 from patent
gb_pat: I26725	+ 28.00	114.82	284.41	55	I26725 Sequence 13 from patent U
gb_pat: I26726	+ 28.00	114.82	284.41	55	I26726 Sequence 14 from patent U
gb_pat: A45812	+ 27.00	112.68	374.18	44	A45812 Sequence 6 from Patent W
gb_pat: H0MTCVDIFS	+ 27.00	110.38	502.69	59	L32520 Human (clone: 5bp11) T-D
gb_pat: I08597	- 26.00	110.59	489.47	35	I08597 Sequence 5 from Patent W
gb_pat: AR021475	+ 26.00	108.97	602.14	43	AR021475 Sequence 5 from Patent W
gb_pat: AR061313	+ 26.00	108.97	602.14	43	AR061313 Sequence 5 from patent
gb_pat: A05101	+ 26.00	108.61	630.33	45	A05101 Oligonucleotide L14 for s
gb_pat: AR021474	+ 26.00	107.48	729.06	52	AR021474 Sequence 4 from patent
gb_pat: AR061312	+ 26.00	107.48	729.06	52	AR061312 Sequence 4 from patent
gb_in3: TVU07784	+ 26.00	106.62	813.75	58	U07784 Trichomonas vaginalis ATC
em_pat: E10448	- 25.00	111.09	458.98	20	E10448 Primer. 10/1997
gb_pat: E13465	+ 25.00	109.34	574.55	25	E13465 PCR primer for detecting
gb_pat: E13467	+ 25.00	109.34	574.55	25	E13467 PCR primer for detecting
gb_pat: AR063949	- 25.00	107.65	713.42	31	AR063949 Sequence 48 from patent
gb_pat: I92782	- 25.00	107.65	713.42	31	I92782 Sequence 48 from patent U
gb_pat: I50691	+ 25.00	106.92	782.93	34	I50691 Sequence 12 from patent U
gb_pat: AR007184	+ 25.00	106.48	829.28	36	AR007184 Sequence 2 from patent
gb_pat: AR046087	+ 25.00	106.48	829.28	36	AR046087 Sequence 2 from patent
gb_pat: AR054191	+ 25.00	106.48	829.28	36	AR054191 Sequence 5 from patent
gb_pat: AR063951	+ 25.00	106.48	829.28	36	AR063951 Sequence 6 from patent
gb_pat: I74400	+ 25.00	106.48	829.28	36	I74400 Sequence 2 from patent US
gb_pat: I43205	- 25.00	105.46	945.25	41	I43205 Sequence 23 from patent U
gb_pat: AR032696	- 25.00	104.55	1.1e+03	46	AR032696 Sequence 308 from patent
gb_pat: I29436	- 25.00	104.55	1.1e+03	46	I29436 Sequence 308 from patent
gb_pat: I91110	- 25.00	104.55	1.1e+03	46	I91110 Sequence 308 from patent
gb_pat: A61802	+ 25.00	104.38	1.1e+03	47	A61802 Sequence 25 from Patent W
gb_ov: EU75790	+ 25.00	103.01	1.3e+03	56	U75790 Raja eglanteria T cell an
gb_pat: A11148	- 25.00	102.87	1.3e+03	57	A11148 Sequence 45 from Patent W
gb_ov: H8ANWSB	+ 25.00	102.87	1.3e+03	57	X72095 H.sapiens (patient Anw, c
gb_ov: EU75789	+ 25.00	102.47	1.4e+03	60	U75789 Raja eglanteria T cell an
gb_in3: DMWST84DA	+ 25.00	102.47	1.4e+03	60	X78359 D.melanogaster Mst84Da le
gb_pat: I46730	- 24.00	107.60	717.87	19	I46730 Sequence 709 from patent

gb_pat: A26000	24.00	105.77	908.13	24	A26000 Artificial DNA for oli
gb_pat: AR052078	24.00	105.77	908.13	24	AR052078 Sequence 48 from pat
gb_pat: I86313	24.00	105.77	908.13	24	I86313 Sequence 48 from paten
gb_pat: I33828	24.00	104.84	1.0e+03	27	I33828 Sequence 22 from paten
seq_name: gb_pat: I03346					
seq_documentation_block:	36 bp ss-DNA				
LOCUS	I03346				PAT
DEFINITION	Sequence 4 from Patent US 4885252.				
ACCESSION	I03346				
VERSION	I03346.1	GI:270672			
KEYWORDS	Unknown.				
SOURCE	Unknown.				
ORGANISM	Unclassified.				
REFERENCE	1 (bases 1 to 36)				
AUTHORS	Ingolia, T.B., Queener, S.W., Skatrud, P.L. and Wlegel, B.J.				
TITLE	Recombinant DNA expression vectors and DNA compounds that encode isoprenicillin N synthetase from aspergillus nidulans				
JOURNAL	Patent: US 4885252-A 4 05-DEC-1989;				
	Eli Lilly and Company;				
	Indianapolis, IN				
FEATURES	Location/Qualifiers				
source	1..36				
BASE COUNT	12 a 8 c 8 g 8 t				
ORIGIN	/organism="unknown"				
alignment_scores:	Quality: 29.00 Length: 8				
	Ratio: 3.625 Gaps: 0				
Percent Similarity:	100.000 Percent Identity: 75.000				
alignment_block:					
US-09-277-074-10 x I03346/rev ..					
Align seg 1/1 to reverse of: I03346 from: 1 to: 36					
2 IlePheGlySerLeuAlaPheLeu 9					
35 ATCTTGGACATGCTTGCTG 12					
seq_name: gb_pat: A82190					
seq_documentation_block:	30 bp DNA				
LOCUS	A82190				PAT
DEFINITION	Sequence 11 from Patent WO9856919.				
ACCESSION	A82190				
VERSION	A82190.1	GI:6732050			
KEYWORDS	unidentified.				
SOURCE	unidentified.				
ORGANISM	unclassified.				
REFERENCE	1 (bases 1 to 30)				
AUTHORS	Gilbert, S.C. and Hill, A.V.				
TITLE	METHODS AND REAGENTS FOR VACCINATION WHICH GENERATE A CD8 T CELL IMMUNE RESPONSE				
JOURNAL	Patent: WO 9856919-A 17-DEC-1998;				
	GILBERT SARAH CATHERINE (GB); HILL ADRIAN VIVIAN SINTON (GB)				
FEATURES	Location/Qualifiers				
source	1..30				
	/organism="unidentified"				
	/db_xref="taxon:32644"				
CDS	1..>30				
	/note="unnamed protein product"				
	/codon_start=1				
	/protein_id="CAB69213.1"				
	/db_xref="GI:6732051"				
	/translation="MNPDPNPNV"				
BASE COUNT	13 a 9 c 4 g 4 t				
ORIGIN					

alignment\_scores:  
 Quality: 28.00 Length: 7  
 Ratio: 4.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 71.429

## alignment\_block:

US-09-277-074-10 x A82190/rev ..  
 Align seg 1/1 to reverse of: A82190 from: 1 to: 30

2 IlePheGlySerLeuAlaPhe 8  
 :::::::::::::::::::::  
 23 CTGTTGGTCATTAGGTTTC 3

seq\_name: gb\_pat:120887

seq\_documentation\_block: 47 bp DNA PAT 07-OCT-1996

LOCUS 120887 Sequence 10 from patent US 5516678.

DEFINITION 120887

ACCESSION 120887

VERSION 120887.1 GI:1601242

KEYWORDS

SOURCE

ORGANISM

Unclassified.

REFERENCE 1 (bases 1 to 47)

AUTHORS Benner, J.S. and Coe, L.H.

TITLE Method for producing the SSPI restriction endonuclease and

methylese

JOURNAL Patent: US 5516678-A 10 14-MAY-1996;

FEATURES

Location/Qualifiers

source

1..47

/organism="unknown"

BASE COUNT 9 a 8 c 12 g 18 t

ORIGIN

alignment\_scores:  
 Quality: 28.00 Length: 7  
 Ratio: 4.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 71.429

## alignment\_block:

US-09-277-074-10 x 120887 ..  
 Align seg 1/1 to: 120887 from: 1 to: 47

2 IlePheGlySerLeuAlaPhe 8  
 :::::::::::::::::::::  
 22 ATATTGGTACCTTGAGTTTC 42

seq\_name: gb\_pat:AR032803

seq\_documentation\_block: 50 bp DNA PAT 29-SEP-1999

LOCUS AR032803 Sequence 415 from patent US 5869241.

DEFINITION AR032803

ACCESSION AR032803

VERSION AR032803.1 GI:5948408

KEYWORDS

SOURCE

ORGANISM

Unclassified.

REFERENCE 1 (bases 1 to 50)

AUTHORS Edwards, C.A., Cantor, C.R., Andrews, B.M., Turin, L.M. and Fry, K.E.

TITLE Method of determining DNA sequence preference of a DNA-binding

molecule

JOURNAL Patent: US 5869241-A 415 09-FEB-1999;

FEATURES

Location/Qualifiers

source

1..50

/organism="unknown"

BASE COUNT 26 a 5 c 8 g 11 t

ORIGIN

## ORIGIN

alignment\_scores:  
 Quality: 28.00 Length: 8  
 Ratio: 4.667 Gaps: 0  
 Percent Similarity: 75.000 Percent Identity: 75.000

## alignment\_block:

US-09-277-074-10 x AR032803/rev ..  
 Align seg 1/1 to reverse of: AR032803 from: 1 to: 50

2 IlePheGlySerLeuAlaPheLeu 9  
 :::::::::::::::::::::  
 50 ATCTTTGGTAGCACAAATTTTA 27

seq\_name: gb\_pat:129543

seq\_documentation\_block: 50 bp DNA PAT 06-FEB-1997

LOCUS 129543 Sequence 415 from patent US 5578444.

DEFINITION 129543

ACCESSION 129543

VERSION 129543.1 GI:1820334

KEYWORDS

SOURCE

ORGANISM

Unknown.

Unclassified.

REFERENCE 1 (bases 1 to 50)

AUTHORS Edwards, C.A., Cantor, C.R., Andrews, B.M., Turin, L.M. and Fry, K.E.

TITLE Sequence-directed DNA-binding molecules compositions and methods

JOURNAL Patent: US 5578444-A 415 26-NOV-1996;

FEATURES

Location/Qualifiers

source

1..50

/organism="unknown"

BASE COUNT 26 a 5 c 8 g 11 t

ORIGIN

alignment\_scores:  
 Quality: 28.00 Length: 8  
 Ratio: 4.667 Gaps: 0  
 Percent Similarity: 75.000 Percent Identity: 75.000

## alignment\_block:

US-09-277-074-10 x 129543/rev ..  
 Align seg 1/1 to reverse of: 129543 from: 1 to: 50

2 IlePheGlySerLeuAlaPheLeu 9  
 :::::::::::::::::::::  
 50 ATCTTTGGTAGCACAAATTTTA 27

seq\_name: gb\_pat:191217

seq\_documentation\_block: 50 bp DNA PAT 01-DEC-1998

LOCUS 191217 Sequence 415 from patent US 5726014.

DEFINITION 191217

ACCESSION 191217

VERSION 191217.1 GI:3935687

KEYWORDS

SOURCE

ORGANISM

Unknown.

Unclassified.

REFERENCE 1 (bases 1 to 50)

AUTHORS Edwards, C.A., Cantor, C.R., Andrews, B.M. and Turin, L.M.

TITLE Screening assay for the detection of DNA-binding molecules

JOURNAL Patent: US 5726014-A 415 10-MAR-1998;

FEATURES

Location/Qualifiers

source

1..50

/organism="unknown"

BASE COUNT 26 a 5 c 8 g 11 t

ORIGIN

alignment\_scores:  
 Quality: 28.00 Length: 8  
 Ratio: 4.667 Gaps: 0  
 Percent Similarity: 75.000 Percent Identity: 75.000

alignment\_block:  
 US-09-277-074-10 x I91217/rev ..  
 Align seg 1/1 to reverse of: I91217 from: 1 to: 50

2 IlePheGlySerLeuAlaPheLeu 9  
 |||||  
 50 ATCTTGTTCACACAAATTTT 27

seq\_name: gb\_pat:AR014596

seq\_documentation\_block: PAT 05-DEC-1998  
 LOCUS AR014596 55 bp DNA  
 DEFINITION Sequence 20 from patent US 5773691.  
 ACCESSION AR014596  
 VERSION AR014596.1 GI:3972050  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unknown.  
 UNCLASSIFIED.  
 REFERENCE 1 (bases 1 to 55)  
 AUTHORS Falco,S.C., Keeler,S.Jo and Rice,J.Ann.  
 TITLE Chimeric genes and methods for increasing the lysine and threonine content of the seeds of plants  
 JOURNAL Patent: US 5773691-A 20 30-JUN-1998;  
 FEATURES Location/Qualifiers  
 source 1..55  
 BASE COUNT 20 a 5 c 23 g 7 t  
 ORIGIN

alignment\_scores:  
 Quality: 28.00 Length: 8  
 Ratio: 3.500 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 62.500

alignment\_block:  
 US-09-277-074-10 x AR014596/rev ..  
 Align seg 1/1 to reverse of: AR014596 from: 1 to: 55

2 IlePheGlySerLeuAlaPheLeu 9  
 |||||  
 36 ATCTTCTTCATCGCCTTCATC 13

seq\_name: gb\_pat:AR014597

seq\_documentation\_block: PAT 05-DEC-1998  
 LOCUS AR014597 55 bp DNA  
 DEFINITION Sequence 21 from patent US 5773691.  
 ACCESSION AR014597  
 VERSION AR014597.1 GI:3972051  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unknown.  
 UNCLASSIFIED.  
 REFERENCE 1 (bases 1 to 55)  
 AUTHORS Falco,S.C., Keeler,S.Jo and Rice,J.Ann.  
 TITLE Chimeric genes and methods for increasing the lysine and threonine content of the seeds of plants  
 JOURNAL Patent: US 5773691-A 21 30-JUN-1998;  
 FEATURES Location/Qualifiers  
 source 1..55  
 BASE COUNT 8 a 22 c 4 g 21 t  
 ORIGIN

ORIGIN

alignment\_scores:  
 Quality: 28.00 Length: 8  
 Ratio: 3.500 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 62.500

alignment\_block:  
 US-09-277-074-10 x AR014597 ..  
 Align seg 1/1 to: AR014597 from: 1 to: 55

2 IlePheGlySerLeuAlaPheLeu 9  
 |||||  
 24 ATCTTCTTCATCGCCTTCATC 47

seq\_name: gb\_pat:I26725

seq\_documentation\_block: PAT 07-OCT-1996  
 LOCUS I26725 55 bp DNA  
 DEFINITION Sequence 13 from patent US 5559223.  
 ACCESSION I26725  
 VERSION I26725.1 GI:1606595  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unknown.  
 UNCLASSIFIED.  
 REFERENCE 1 (bases 1 to 55)  
 AUTHORS Falco,S.C., Keeler,S.J. and Rice,J.A.  
 TITLE Synthetic storage proteins with defined structure containing programmable levels of essential amino acids for improvement of the nutritional value of plants  
 JOURNAL Patent: US 5559223-A 13 24-SEP-1996;  
 FEATURES Location/Qualifiers  
 source 1..55  
 BASE COUNT 20 a 5 c 23 g 7 t  
 ORIGIN

alignment\_scores:  
 Quality: 28.00 Length: 8  
 Ratio: 3.500 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 62.500

alignment\_block:  
 US-09-277-074-10 x I26725/rev ..  
 Align seg 1/1 to reverse of: I26725 from: 1 to: 55

2 IlePheGlySerLeuAlaPheLeu 9  
 |||||  
 36 ATCTTCTTCATCGCCTTCATC 13

seq\_name: gb\_pat:I26726

seq\_documentation\_block: PAT 07-OCT-1996  
 LOCUS I26726 55 bp DNA  
 DEFINITION Sequence 14 from patent US 5559223.  
 ACCESSION I26726  
 VERSION I26726.1 GI:1606596  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unknown.  
 UNCLASSIFIED.  
 REFERENCE 1 (bases 1 to 55)  
 AUTHORS Falco,S.C., Keeler,S.J. and Rice,J.A.  
 TITLE Synthetic storage proteins with defined structure containing programmable levels of essential amino acids for improvement of the nutritional value of plants  
 JOURNAL Patent: US 5559223-A 14 24-SEP-1996;  
 FEATURES Location/Qualifiers

```
source 1..55
/organism="unknown"
BASE COUNT 8 a 22 c 4 g 21 t
ORIGIN

alignment_scores:
  Quality: 28.00 Length: 8
  Ratio: 3.500 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 62.500

alignment_block:
US-09-277-074-10 x I26726 ..
Align seg 1/1 to: I26726 from: 1 to: 55
2 IlePheGlySerLeuAlaPheLeu 9
|||||:|||||:|||||:
24 ATCTTCTTCATCGCCTTCATC 47

seq_name: gb_pat:A45812
seq_documentation_block:
LOCUS A45812 44 bp DNA PAT
DEFINITION Sequence 6 from Patent WO9519371.
ACCESSION A45812
VERSION A45812.1 GI:2300185
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 44)
AUTHORS Deblaeere, R.Y., Desomer, J. and Dhaese, P.
TITLE EXPRESSION OF SURFACE LAYER PROTEINS
JOURNAL Patent: WO 9519371-A 6 20-JUL-1995;
SOLVAY. (BE)
FEATURES
  source
  location/Qualifiers
BASE COUNT 16 a 5 c 6 g 17 t
ORIGIN

alignment_scores:
  Quality: 27.00 Length: 9
  Ratio: 3.857 Gaps: 0
Percent Similarity: 77.778 Percent Identity: 66.667

alignment_block:
US-09-277-074-10 x A45812 ..
Align seg 1/1 to: A45812 from: 1 to: 44
1 LysilePheGlySerLeuAlaPheLeu 9
|||||:|||||:|||||:
2 AAATATACGGAGCTCTTAATTTTG 28

seq_name: gb_pr7:HUMTCVDIFS
seq_documentation_block:
LOCUS HUMTCVDIFS 59 bp mRNA PRI 10-FEB-1995
DEFINITION Human (clone: 5ppl1) T-cell receptor delta-chain (V-delta-1) mRNA.
ACCESSION L32520
VERSION L32520.1 GI:497590
KEYWORDS T-cell receptor; delta chain.
SOURCE Homo sapiens peripheral blood cDNA to mRNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 59)
AUTHORS Chowdhury, Y., Holtmeier, W., Harwood, J., Morzycka-Wroblewska, E. and
Kagnoff, M.F.

source 1..55
/organism="unknown"
BASE COUNT 8 a 22 c 4 g 21 t
ORIGIN

alignment_scores:
  Quality: 28.00 Length: 8
  Ratio: 3.500 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 62.500

alignment_block:
US-09-277-074-10 x I26726 ..
Align seg 1/1 to: I26726 from: 1 to: 55
2 IlePheGlySerLeuAlaPheLeu 9
|||||:|||||:|||||:
24 ATCTTCTTCATCGCCTTCATC 47

seq_name: gb_pat:A45812
seq_documentation_block:
LOCUS A45812 44 bp DNA PAT
DEFINITION Sequence 6 from Patent WO9519371.
ACCESSION A45812
VERSION A45812.1 GI:2300185
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 44)
AUTHORS Deblaeere, R.Y., Desomer, J. and Dhaese, P.
TITLE EXPRESSION OF SURFACE LAYER PROTEINS
JOURNAL Patent: WO 9519371-A 6 20-JUL-1995;
SOLVAY. (BE)
FEATURES
  source
  location/Qualifiers
BASE COUNT 16 a 5 c 6 g 17 t
ORIGIN

alignment_scores:
  Quality: 27.00 Length: 9
  Ratio: 3.857 Gaps: 0
Percent Similarity: 77.778 Percent Identity: 66.667

alignment_block:
US-09-277-074-10 x A45812 ..
Align seg 1/1 to: A45812 from: 1 to: 44
1 LysilePheGlySerLeuAlaPheLeu 9
|||||:|||||:|||||:
2 AAATATACGGAGCTCTTAATTTTG 28

seq_name: gb_pr7:HUMTCVDIFS
seq_documentation_block:
LOCUS HUMTCVDIFS 59 bp mRNA PRI 10-FEB-1995
DEFINITION Human (clone: 5ppl1) T-cell receptor delta-chain (V-delta-1) mRNA.
ACCESSION L32520
VERSION L32520.1 GI:497590
KEYWORDS T-cell receptor; delta chain.
SOURCE Homo sapiens peripheral blood cDNA to mRNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 59)
AUTHORS Chowdhury, Y., Holtmeier, W., Harwood, J., Morzycka-Wroblewska, E. and
Kagnoff, M.F.

TITLE The V delta 1 T cell receptor repertoire in human small intestine
and colon
JOURNAL J. Exp. Med. 180 (1), 183-190 (1994)
MEDLINE 94275371
FEATURES
  source
  location/Qualifiers
  organism="Homo sapiens"
  db_xref="taxon:9606"
  tissue_type="peripheral blood"
BASE COUNT 12 a 18 c 14 g 15 t
ORIGIN

alignment_scores:
  Quality: 27.00 Length: 8
  Ratio: 3.375 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 62.500

alignment_block:
US-09-277-074-10 x HUMTCVDIFS ..
Align seg 1/1 to: HUMTCVDIFS from: 1 to: 59
2 IlePheGlySerLeuAlaPheLeu 9
|||||:|||||:|||||:
1 CTCITGGGGACCTTGGCCTTCCTA 24

seq_name: gb_pat:I08597
seq_documentation_block:
LOCUS I08597 35 bp PAT
DEFINITION Sequence 12 from Patent WO 8707144.
ACCESSION I08597
VERSION I08597.1 GI:588701
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 35)
AUTHORS Kaufman, R.J., Pittman, D.D. and Toole, J.J.
TITLE NOVEL PROCOAGULANT PROTEINS
JOURNAL Patent: WO 8707144-A 12 03-DEC-1987;
FEATURES
  source
  location/Qualifiers
BASE COUNT 18 a 8 c 5 g 4 t
ORIGIN

alignment_scores:
  Quality: 26.00 Length: 8
  Ratio: 4.333 Gaps: 0
Percent Similarity: 75.000 Percent Identity: 62.500

alignment_block:
US-09-277-074-10 x I08597/rev ..
Align seg 1/1 to reverse of: I08597 from: 1 to: 35
1 LysilePheGlySerLeuAlaPhe 8
|||||:|||||:|||||:
24 AAGCTTTTGGGGCTCTGATTTTC 1

seq_name: gb_pat:AR021475
seq_documentation_block:
LOCUS AR021475 43 bp DNA PAT
DEFINITION Sequence 5 from patent US 5789651.
ACCESSION AR021475
VERSION AR021475.1 GI:3976090
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
```

```
Unclassified.
REFERENCE 1 (bases 1 to 43)
AUTHORS Koychik,R.P.
TITLE Isolation and characterization of Agouti: a diabetes/obesity
related gene
JOURNAL Patent: US 5789651-A 5 04-AUG-1998;
FEATURES Location/Qualifiers
source 1..43
/organism="unknown"
BASE COUNT 9 a 5 c 12 g 17 t
ORIGIN

alignment_scores:
Quality: 26.00 Length: 9
Ratio: 2.889 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 55.556

alignment_block:
US-09-277-074-10 x AR021475 ..
Align seg 1/1 to: AR021475 from: 1 to: 43

1 LysilePheGlySerLeuAlaPheLeu 9
|||||:|||||:|||||:|||||
6 AAGATGCTTGGTGGACTGGTTTCTT 32

seq_name: gb_pat:AR061313

seq_documentation_block:
LOCUS AR061313 43 bp DNA PAT 29-SEP-1999
DEFINITION Sequence 5 from patent US 5843652.
ACCESSION AR061313
VERSION AR061313.1 GI:5989004
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 43)
AUTHORS Koychik,R.P.
TITLE Isolation and characterization of Agouti: a diabetes/obesity
related gene
JOURNAL Patent: US 5843652-A 5 01-DEC-1998;
FEATURES Location/Qualifiers
source 1..43
/organism="unknown"
BASE COUNT 9 a 5 c 12 g 17 t
ORIGIN

alignment_scores:
Quality: 26.00 Length: 9
Ratio: 2.889 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 55.556

alignment_block:
US-09-277-074-10 x AR061313 ..
Align seg 1/1 to: AR061313 from: 1 to: 43

1 LysilePheGlySerLeuAlaPheLeu 9
|||||:|||||:|||||:|||||
6 AAGATGCTTGGTGGACTGGTTTCTT 32
```



OM of: US-09-277-074-10 to: N\_Geneseq\_36.\* out\_format : pfs

Date: Nov 16, 2000 1:08 AM

About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 Compugen Ltd.

# Command line parameters:

-MODEL-frame+pn2n.model -DEV-xlp  
-Q/cn2\_1/USPTO\_spool/US09277074/runat\_14112000\_120417\_23101/app\_query.fasta\_1.67  
-DB-N\_Geneseq\_36 -QFMT-fastaf -SUFFIX-lim60.rng -GAPOP-12.000  
-GAPEXT-4.000 -MINMATCH-0.100 -LOOPCL-0.000 -LOOPEXT-0.000  
-GAPOP-4.500 -OGAPEXT-0.050 -XGAPOP-10.000 -XGAPEXT-0.500  
-FGAPOP-6.000 -FGAPEXT-7.000 -YGAPOP-10.000 -YGAPEXT-0.500  
-DELOP-6.000 -DELEXT-7.000 -START-1 -MATRIX-biosum62  
-TRANS-human40.cdi -LIST-45 -DOCALLIGN-200 -THR\_SCORE-pct  
-THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE-LOCAL -OUTFMT-pfs  
-NORM-ext -MINLEN=0 -MAXLEN=60 -USER-US09277074\_@CGN1\_1\_108  
-NCPU=6 -ICPU=3 -LONGLOG -NO\_XLPXY -WAIT -THREADS=1

# Search information block:

Query: US-09-277-074-10  
Query length: 9  
Database: N\_Geneseq\_36.\*  
Database sequences: 480022  
Database length: 187831343  
Search time (sec): 79.960000

# score\_list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
/SID56/gcgdata/geneseq/geneseq/NA1997.DAT:T74842	-	30.00	130.30	18.09	24
/SID56/gcgdata/geneseq/geneseq/NA1997.DAT:T74843	+	30.00	130.30	18.09	24
/SID56/gcgdata/geneseq/geneseq/NA1998.DAT:T50805	+	29.00	121.89	53.19	41
/SID56/gcgdata/geneseq/geneseq/NA1999.DAT:X29206	+	28.00	120.83	60.94	30
/SID56/gcgdata/geneseq/geneseq/NA1996.DAT:T13501	+	28.00	116.93	100.40	47
/SID56/gcgdata/geneseq/geneseq/NA1994.DAT:Q69665	+	28.00	116.40	107.55	50
/SID56/gcgdata/geneseq/geneseq/NA1997.DAT:T64117	+	28.00	116.40	107.55	50
/SID56/gcgdata/geneseq/geneseq/NA1999.DAT:X17415	-	28.00	116.40	107.55	50
/SID56/gcgdata/geneseq/geneseq/NA1993.DAT:Q36814	+	28.00	115.57	119.57	55
/SID56/gcgdata/geneseq/geneseq/NA1993.DAT:Q36815	+	28.00	115.57	119.57	55
/SID56/gcgdata/geneseq/geneseq/NA1995.DAT:Q94972	-	28.00	115.57	119.57	55
/SID56/gcgdata/geneseq/geneseq/NA1995.DAT:Q94973	+	28.00	115.57	119.57	55
/SID56/gcgdata/geneseq/geneseq/NA1999.DAT:T95059	+	28.00	115.57	119.57	55
/SID56/gcgdata/geneseq/geneseq/NA1994.DAT:Q63720	+	28.00	115.42	121.99	56
/SID56/gcgdata/geneseq/geneseq/NA1997.DAT:T68905	+	27.00	117.66	91.52	28
/SID56/gcgdata/geneseq/geneseq/NA1999.DAT:X22057	+	27.00	115.48	121.03	36
/SID56/gcgdata/geneseq/geneseq/NA1999.DAT:X22058	+	27.00	115.24	124.77	37
/SID56/gcgdata/geneseq/geneseq/NA1995.DAT:Q99427	+	27.00	113.74	131.28	44
/SID56/gcgdata/geneseq/geneseq/NA1995.DAT:T25609	+	27.00	111.34	205.67	58
/SID56/gcgdata/geneseq/geneseq/NA1998.DAT:T72751	+	26.00	115.98	113.50	22
/SID56/gcgdata/geneseq/geneseq/NA1999.DAT:X22034	+	26.00	114.87	130.83	25
/SID56/gcgdata/geneseq/geneseq/NA1998.DAT:T98093	-	26.00	114.20	142.52	27
/SID56/gcgdata/geneseq/geneseq/NA1998.DAT:T98857	-	26.00	114.20	142.52	27
/SID56/gcgdata/geneseq/geneseq/NA1997.DAT:T79311	-	26.00	113.00	166.18	31
/SID56/gcgdata/geneseq/geneseq/NA2000.DAT:A14362	+	26.00	112.20	184.15	34
/SID56/gcgdata/geneseq/geneseq/NA1997.DAT:T94012	+	26.00	111.71	196.24	36
/SID56/gcgdata/geneseq/geneseq/NA1999.DAT:X22034	+	26.00	111.71	196.24	36
/SID56/gcgdata/geneseq/geneseq/NA1998.DAT:T42358	+	26.00	110.17	239.10	43
/SID56/gcgdata/geneseq/geneseq/NA1999.DAT:T42358	+	26.00	110.17	239.10	43
/SID56/gcgdata/geneseq/geneseq/NA1998.DAT:N00698	-	26.00	109.77	221.49	45
/SID56/gcgdata/geneseq/geneseq/NA1998.DAT:T42357	+	26.00	108.52	295.35	52
/SID56/gcgdata/geneseq/geneseq/NA1998.DAT:T42357	+	26.00	108.52	295.35	52
/SID56/gcgdata/geneseq/geneseq/NA1995.DAT:T26751	-	26.00	108.36	301.67	53
/SID56/gcgdata/geneseq/geneseq/NA1998.DAT:T54157	-	25.00	113.48	156.35	19
/SID56/gcgdata/geneseq/geneseq/NA2000.DAT:T36528	+	25.00	113.48	156.35	19
/SID56/gcgdata/geneseq/geneseq/NA1996.DAT:T16262	+	25.00	112.61	174.75	20
/SID56/gcgdata/geneseq/geneseq/NA1999.DAT:X35983	+	25.00	112.61	174.75	21
/SID56/gcgdata/geneseq/geneseq/NA1995.DAT:Q90191	+	25.00	111.10	212.13	25
/SID56/gcgdata/geneseq/geneseq/NA1997.DAT:V00191	+	25.00	111.10	212.13	25
/SID56/gcgdata/geneseq/geneseq/NA1997.DAT:V00189	+	25.00	111.10	212.13	25
/SID56/gcgdata/geneseq/geneseq/NA1999.DAT:T12371	-	25.00	110.76	221.59	26

/SID56/gcgdata/geneseq/geneseq/NA1998.DAT:V14045 + 25.00 110.43 231.09  
/SID56/gcgdata/geneseq/geneseq/NA1997.DAT:T59549 + 25.00 109.81 250.20  
/SID56/gcgdata/geneseq/geneseq/NA1997.DAT:T59563 + 25.00 109.81 250.20  
/SID56/gcgdata/geneseq/geneseq/NA1993.DAT:Q38549 - 25.00 109.52 259.81

seq\_name: /SID56/gcgdata/geneseq/geneseq/NA1997.DAT:T74842

# seq\_documentation block:

ID T74842 standard; cDNA; 24 BP.  
XX AC T74842;  
XX DT 10-FEB-1998 (first entry)  
XX DE Porcine retrovirus PCR primer #29.  
XX KW Retrovirus; porcine; GAG protein; POL protein; ENV protein;  
KW xenotransplantation; infectious; provirus; organ transplant; donor;  
KW activated virus; Tsukuba-1; PCR; primer; ss.  
XX OS Synthetic.  
OS Porcine retrovirus.  
XX PN WO9721836-A1.  
XX PD 19-JUN-1997.  
XX PF 13-DEC-1996; 96WO-US19680.  
XX PR 14-DEC-1995; 95US-0572645.  
XX PA (GEHO) GEN HOSPITAL CORP.

XX PI Fishman JA;

XX DR WPI; 1997-332804/30.

XX PT New nucleic acid from porcine retroviruses - used for detecting

XX PT viruses in transplant or other tissue and for assessing risk of

XX PT transmitting infection to graft recipient

XX PS Claim 10; Page 68; 128pp; English.

XX CC This PCR primer is designed from one of the following sequences: the  
XX CC porcine retrovirus Tsukuba-1 cDNA, the genome of a defective porcine  
XX CC retrovirus found in PK-15 cells and a retrovirus from miniature swine.  
XX CC Fragments generated from the amplification of such viral sequences as  
XX CC the GAG, POL and ENV viral proteins could be used to screen organs for  
XX CC porcine retroviruses prior to xenotransplantation. Transplantation can  
XX CC increase the likelihood of retroviral activation if intact and infectious  
XX CC proviruses are present. The porcine retroviral sequence can be used to  
XX CC generate probes to determine the level (e.g. copy number) of intact  
XX CC (i.e. potentially replicating) porcine provirus sequences in a strain of  
XX CC xenograft transplantation donors. It can be used to detect mutations,  
XX CC genetic lesions or viral recombinants and to determine the histological  
XX CC localisation of activated retrovirus. Using Polymerase Chain Reaction  
XX CC DNA Quantitation (PDQ) on blood mononuclear cells, infectivity titration  
XX CC and susceptibility testing can be performed. Ultimately animal donors  
XX CC without intact porcine retroviral sequences or with a lower copy number  
XX CC of viral elements could be selected.

SQ Sequence 24 BP; 9 A; 8 C; 6 G; 1 T; 0 other;

alignment\_scores:  
Quality: 30.00 Length: 7  
Ratio: 4.286 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 85.714

# alignment\_block:

US-09-277-074-10 x T74842/rev ..  
Align seg 1/1 to reverse of: T74842 from: 1 to: 24

3 PheGlySerLeuAlapheteu 9  
|||||:|||||  
24 TTCGGGAGCTGGCTTCTTG 4

seq\_name: /SID56/gcgdata/geneseq/geneseqn/NA1997.DAT:T74843

seq\_documentation\_block:  
ID T74843 standard; CDNA; 24 BP.

XX AC T74843;  
XX DT 10-FEB-1998 (first entry)  
XX DE Porcine retrovirus PCR primer #30.

XX KW Retrovirus; porcine; GAG protein; POL protein; ENV protein;  
XX KW xenotransplantation; infectious; provirus; organ transplant; donor;  
XX KW activated virus; Tsukuba-1; PCR; primer; ss.

XX OS Synthetic.  
XX OS Porcine retrovirus.

XX PN WO9721836-A1.

XX PD 19-JUN-1997.

XX PF 13-DEC-1996; 96WO-US19680.

XX PR 14-DEC-1995; 95US-0572645.

XX PA (GEO) GEN HOSPITAL CORP.

XX PI Fishman JA;

XX DR WPI; 1997-332804/30.

XX PT New nucleic acid from porcine retroviruses - used for detecting  
XX PT viruses in transplant or other tissue and for assessing risk of  
XX PT transmitting infection to graft recipient

XX PS Claim 10; Page 69; 128pp; English.

XX CC This PCR primer is designed from one of the following sequences; the  
XX CC porcine retrovirus Tsukuba-1 cDNA, the genome of a defective porcine  
XX CC retrovirus found in PK-15 cells and a retrovirus from miniature swine.  
XX CC Fragments generated from the amplification of such viral sequences as  
XX CC the GAG, POL and ENV viral proteins could be used to screen organs for  
XX CC porcine retroviruses prior to xenotransplantation. Transplantation can  
XX CC increase the likelihood of retroviral activation if intact and infectious  
XX CC proviruses are present. The porcine retroviral sequence can be used to  
XX CC generate probes to determine the level (e.g. copy number) of intact  
XX CC (i.e. potentially replicating) porcine provirus sequences in a strain of  
XX CC xenograft transplantation donors. It can be used to detect mutations,  
XX CC genetic lesions or viral recombinants and to determine the histological  
XX CC localisation of activated retrovirus. Using Polymerase Chain Reaction  
XX CC DNA Quantitation (PQO) on blood mononuclear cells, infectivity titration  
XX CC and susceptibility testing can be performed. Ultimately animal donors  
XX CC without intact porcine retroviral sequences or with a lower copy number  
XX CC of viral elements could be selected.

XX SQ Sequence 24 BP; 1 A; 6 C; 8 G; 9 T; 0 other;

alignment\_scores:  
Quality: 30.00 Length: 7  
Ratio: 4.286 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 85.714

alignment\_block:  
US-09-277-074-10 x T74843 ..

Align seq 1/1 to: T74843 from: 1 to: 24

3 PheGlySerLeuAlapheteu 9  
|||||:|||||  
1 TTCGGGAGCTGGCTTCTTG 21

seq\_name: /SID56/gcgdata/geneseq/geneseqn/NA1998.DAT:V50805

seq\_documentation\_block:  
ID V50805 standard; DNA; 41 BP.

XX AC V50805;

XX DT 04-JAN-1999 (first entry)

XX DE Brassica sp. polymorphic marker 85/20D6/86-5 DNA.

XX KW Polymorphic marker; allele-specific; primer; probe; amplification;  
XX KW hybridisation; plant; hybrid certification; genetic contribution;  
XX KW progeny; back-cross; hybrid; ancestry; ss.

XX OS Brassica sp.

XX FH Key Location/Qualifiers  
XX FT variation 21..22

XX FT /\*tag= a

XX FT /replace= "agg"

XX FT /note= "polymorphism"

XX PN WO9824796-A1.

XX PD 11-JUN-1998.

XX PF 01-DEC-1997; 97WO-US21782.

XX PR 07-MAR-1997; 97US-0813507.

XX PR 02-DEC-1996; 96US-0032069.

XX PA (AFFY-) AFFYMETRIX INC.

XX PI Landry BS, Lemieux B, Murigneux A, Sapolsky RJ;

XX DR WPI; 1998-33252/29.

XX PT Brassica species allele-specific oligonucleotide probes and primers

XX FT - useful for plant breeding

XX PS Claim 1; Page 39; 65pp; English.

XX CC This DNA sequence is a region of a Brassica napus or Brassica oleracea  
XX CC genome which contains a polymorphic marker. This sequence can be used  
XX CC in the construction of allele-specific primers and probes for  
XX CC amplification or hybridisation, e.g. to determine common or disparate  
XX CC ancestry between 2 or more plants, to monitor the genetic contribution  
XX CC of an ancestral plant, to trace the progeny of proprietary plants, in  
XX CC certification of a hybrid plant or to identify the progeny of a  
XX CC back-crossed plant with an ancestral plant.

XX SQ Sequence 41 BP; 20 A; 10 C; 7 G; 4 T; 0 other;

alignment\_scores:  
Quality: 29.00 Length: 7  
Ratio: 4.143 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 71.429

alignment\_block:  
US-09-277-074-10 x V50805/rev ..

Align seq 1/1 to reverse of: V50805 from: 1 to: 41

1 LysilePheGlySerLeuAla 7  
:|||||:|||||

34 CGAGTGTGGATCTTGGCG 14



seq\_name: /SID56/gcgdata/geneseq/NA1999.DAT:X29206

seq\_documentation\_block:

ID X29206 standard; DNA; 30 BP.

XX AC X29206;

XX DT 07-JUN-1999 (first entry)

XX DE DNA sequence of the malaria (M) string CTL epitope Cp6.

XX KW CD8+ T-cell; immune response; antigen; priming composition; CTL; epitope;  
XX KW cytotoxic T lymphocyte; boosting; poxvirus vector; PVV; pathogen; tumour;  
XX KW malaria; parasite; P. falciparum; viral; bacterial; parasitic; cancer;  
XX KW melanoma; HIV; breast; colon; vaccination; ss.

XX OS Synthetic.

XX OS Plasmodium falciparum.

XX PN WO9856919-A2.

XX PD 17-DEC-1998.

XX PF 09-JUN-1998; 98WO-GB01681.

XX PR 09-JUN-1997; 97GB-0011957.

XX PA (ISIS-) ISIS INNOVATION LTD.

XX PI Blanchard T, Gilbert SC, Hanke T, Hill AVS, McMichael AJ;

XX PI Plebanski M, Schneider J, Smith GL;

XX DR WPI; 1999-070325/06.

XX DR P-PSDB; Y03666.

XX PT Generating CD8-positive T cell response to target antigen using  
XX PT recombinant poxvirus - for treating or preventing malaria and HIV  
XX PT infection, also epitope strings from Plasmodium and HIV

XX PS Example 1; Page 18; 85pp; English.

XX CC The invention relates to methods and reagents for generating a protective  
XX CC CD8+ T-cell immune response against at least one target antigen. The kits  
XX CC of the invention comprises (i) as priming composition, a source of one or  
XX CC more CD8+ T-cell cytotoxic T lymphocytes (CTL) epitopes of the target  
XX CC antigen, plus a carrier and (ii) as boosting composition a source of CTL  
XX CC epitopes, with at least one CTL epitope the same as used in (i), with  
XX CC this source being a non-replicating or replication-impaired recombinant  
XX CC poxvirus vector (PVV) plus a carrier. If the source of CTL epitopes in  
XX CC (i) is a viral vector, then the vector in (ii) is from a different virus.  
XX CC The kits are used to generate an immune response (prophylactic or  
XX CC therapeutic) against pathogens or tumours, specifically against malaria  
XX CC parasites such as P. falciparum, or HIV, and also many other bacterial,  
XX CC viral or parasitic pathogens. The kits are also used for protective  
XX CC response against melanoma and cancer of breast or colon, and generally  
XX CC wherever a strong CD8+ response is protective. The boosting composition  
XX CC may be used alone to boost a naturally primed response against malaria.  
XX CC The specified PVV provide an excellent booster effect, better than that  
XX CC from wild-type poxvirus, resulting in complete rather than partial  
XX CC protection against sporozoite challenge. Also PVV are safer to use than  
XX CC wild-type virus. Sequences X29201-220 represent synthetic DNA sequences  
XX CC of the CTL epitopes of the malaria (M) string.

XX SQ Sequence 30 BP; 13 A; 9 C; 4 G; 4 T; 0 other;

alignment\_scores:

Quality: 28.00 Length: 7  
Ratio: 4.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 71.429

alignment\_block:

US-09-277-074-10 x X29206/rev ..

Align seg 1/1 to reverse of: X29206 from: 1 to: 30

2 IlePheGlySerLeuAlaPhe 8

:::|||||:|||||:|||||

23 CTGTTGGGTCAATAGGGTTC 3

seq\_name: /SID56/gcgdata/geneseq/NA1996.DAT:T13501

seq\_documentation\_block:

ID T13501 standard; DNA; 47 BP.

XX AC T13501;

XX DT 12-JUN-1996 (first entry)

XX DE SspI endonuclease C-terminal PCR primer 136.

XX KW SspI; restriction endonuclease; modification methylase;

XX KW Sphaerotilus; primer; polymerase chain reaction; PCR; ss.

XX OS Synthetic.

XX PN EP707066-A2.

XX PD 17-APR-1996.

XX PF 06-OCT-1995; 95EP-0307106.

XX PR 06-OCT-1994; 94US-0319621.

XX PA (NEWE) NEW ENGLAND BIOLABS INC.

XX XX Benner JS, Coe LH;

XX DR WPI; 1996-189958/20.

XX PT DNA encoding SspI restriction endonuclease and opt. modification  
XX PT methylase, from vector p(PAII17)SspR7.2-B1, also host cell for  
XX PT recombinant prodn. of the enzymes

XX PS Disclosure; Page 13; 46pp; English.

XX CC 2 Primers (T13500-01) were designed for the N-terminal and  
XX CC C-terminal ends, respectively, of the SspI endonuclease of  
XX CC Sphaerotilus sp. ATCC 13925. The primers were used for the PCR  
XX CC amplification of Sphaerotilus sp. genomic DNA. A 900 bp product  
XX CC was cloned into vector pAII17. Vector p(PAII17)SspR7.2-B1 (ATCC  
XX CC 75909) was obtd. This can be used for prodn. of SspI endonuclease  
XX CC (see R94364) in Escherichia coli host cells.

XX SQ Sequence 47 BP; 9 A; 8 C; 12 G; 18 T; 0 other;

alignment\_scores:

Quality: 28.00 Length: 7  
Ratio: 4.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 71.429

alignment\_block:

US-09-277-074-10 x T13501 ..

Align seg 1/1 to: T13501 from: 1 to: 47

2 IlePheGlySerLeuAlaPhe 8

|||||:|||||:|||||

22 ATATTGGTACCTTGAGTTTC 42

seq\_name: /SID56/gcgdata/geneseq/NA1994.DAT:Q69665

seq\_documentation\_block:

ID Q69665 standard; DNA; 50 BP.

```

XX AC Q69665;
XX DT 01-MAR-1995 (first entry)
XX DE Human aldolase B (ALDOB) gene, target region.
XX KW DNA protein-binding assay; test sequence; screening sequence;
XX KW promoter; target; TATA box; Herpes Simplex Virus; HSV;
XX KW origin of replication; UL9; transcription factor; TFIID: ds.
XX OS Synthetic.
XX PN W09414980-A.
XX PD 07-JUL-1994.
XX PF 20-DEC-1993; 93WO-US12388.
XX PR 23-DEC-1992; 92US-0996783.
XX PR 17-SEP-1993; 93US-0123936.
XX PA (GENE-) GENELABS TECHNOLOGIES INC.
XX PI Andrews BM, Cantor CR, Edwards CA, Fry KE, Turin LM;
XX WPI; 1994-234711/28.
XX DR
XX PT Sequence-directed DNA-binding molecules - useful in
XX PT pharmaceuticals and as molecular reagents
XX PS Claim 28; Page 419; 587pp; English.
XX CC A DNA protein-binding assay is provided, useful for screening
XX CC libraries of synthetic or biological cpds. for their ability
XX CC to bind DNA test sequences. The assay is versatile in that any
XX CC number of test sequences can be tested by placing the test sequence
XX CC adjacent to a defined protein-binding screening sequence. Binding
XX CC of mols. to these test sequences changes the binding characteristics
XX CC of the protein mol. to its cognate binding sequence. When such a mol.
XX CC binds the test sequence, the equilibrium of the DNA:protein complexes
XX CC is disturbed, generating changes in the concentration of free DNA probe.
XX CC One application of this method is to eucaryotic general transcription
XX CC factors (e.g. TFIID), where the target region is typically selected
XX CC from DNA sequences adjacent to the binding site for the eucaryotic
XX CC transcription factor. Numerous exemplary test sequences are given:
XX CC the sequences in Q69251-731 and Q69850 correspond to promoter targets
XX CC (typically, TATA box-contg. sites) for human genes and the sequences in
XX CC Q69732-849 correspond to promoter targets for viral genes. The test
XX CC sequences may also be randomly generated. DNA:protein interaction may
XX CC be used for screening purposes, e.g. the Herpes Simplex Virus (HSV)
XX CC origin of replication and UL9 (see Q69851-52, Q69865 and Q69891).
XX SQ Sequence 50 BP; 26 A; 5 C; 8 G; 11 T; 0 other;

alignment_scores:
Quality: 28.00 Length: 8
Ratio: 4.667 Gaps: 0
Percent Similarity: 75.000 Percent Identity: 75.000

alignment_block:
US-09-277-074-10 x Q69665/rev ..
Align seg 1/1 to reverse of: Q69665 from: 1 to: 50
.. 2 IlePheGlySerLeuAlaPheLeu 9
|||||
50 ATCTTGGTAGCACACAAATTTT 27

seq_name: /SBS6/gcgdata/geneseq/geneseq/NA1997.DAT:T64127
seq_documentation_block:

```

```

ID T64127 standard; DNA; 50 BP.
XX AC T64127;
XX DT 17-MAR-1997 (first entry)
XX DE Human aldolase B (ALDOB) gene TFIID binding site.
XX KW Duplex DNA; target region; binding characteristic; DNA binding protein;
XX KW TFIID; transcription factor; binding site; inhibition; enhance;
XX KW cancer; inherited genetic disorder; ds.
XX OS Homo sapiens.
XX PN US5578444-A.
XX PD 26-NOV-1996.
XX PF 27-JUN-1991; 91US-0723618.
XX PR 20-DEC-1993; 93US-0171389.
XX PR 27-JUN-1991; 91US-0723618.
XX PR 23-DEC-1992; 92US-0996783.
XX PR 17-SEP-1993; 93US-0123936.
XX PA (GENE-) GENELABS TECHNOLOGIES INC.
XX PI Andrews BM, Cantor CR, Edwards CA, Fry KE, Turin LM;
XX WPI; 1997-020402/02.
XX DR
XX PT Altering binding characteristics of DNA binding proteins to duplex
XX PT DNA - by attaching specific small cpd. to target region close to the
XX PT protein's binding site, useful in treatment of viral disease, cancer
XX PS Claim 6; Column 311-312; 264pp; English.
XX CC The sequences given in T63713-4312 represent duplex DNA's which act
XX CC as target regions in the method of the invention. The method for
XX CC altering the binding characteristics of a DNA-binding protein to duplex
XX CC DNA comprises contacting the duplex DNA with a small molecule which
XX CC binds sequence-specifically to a target region, where, when the small
XX CC molecule is bound to the target region, it is adjacent to, but not
XX CC overlapping by more than 4 bp, a binding site for a DNA-binding protein.
XX CC The small molecule is added at a concentration effective to alter the
XX CC binding of the DNA binding protein, pref. TFIID, to its binding site on
XX CC the duplex DNA. The binding of the small molecule may inhibit or enhance
XX CC the binding of the DNA-binding protein to its binding site. The
XX CC compounds isolated using this method are potentially useful as
XX CC therapeutic agents for treatment of any disease which involves a
XX CC specific DNA sequence, e.g. cancer, or inherited genetic disorders etc.
XX CC The method is suitable for screening large biological or chemical
XX CC libraries and allows determination of sequence-specific and relative
XX CC affinities of known DNA-binding agents for different DNA sequences.
XX CC The design of these duplex DNA's allows a single DNA:protein interaction
XX CC to be used for screening sequence-specific, or preferential, DNA binding
XX CC proteins that recognise almost any possible sequence (see also T49539-
XX SQ Sequence 50 BP; 26 A; 5 C; 8 G; 11 T; 0 other;

alignment_scores:
Quality: 28.00 Length: 8
Ratio: 4.667 Gaps: 0
Percent Similarity: 75.000 Percent Identity: 75.000

alignment_block:
US-09-277-074-10 x T64127/rev ..
Align seg 1/1 to reverse of: T64127 from: 1 to: 50

```

2 IlePheGlySerLeuAlaPheLeu 9  
|||||  
50 ATCTTGTAGCACACAATTTTA 27

seq\_name: /SIDS6/gcgdata/geneseq/geneseqn/NA1999.DAT.X17415

seq\_documentation\_block:

ID X17415 standard; DNA; 50 BP.

XX

AC X17415;

XX 06-MAY-1999 (first entry)

XX Test sequence from human aldolase B (ALDOB) gene.

DE Test sequence; DNA-binding molecule; screening sequence; human;

XX nucleic acid amplification; target; viral; ds.

XX Homo sapiens.

XX U55869241-A.

XX 09-FEB-1999.

XX 07-JUN-1995; 95US-0475228.

XX 20-DEC-1993; 93US-0171389.

XX 27-JUN-1991; 91US-0723818.

XX 23-DEC-1992; 92US-0996783.

XX 17-SEP-1993; 93US-0123936.

XX 07-JUN-1995; 95US-0475228.

XX (GENE-) GENELABS TECHNOLOGIES INC.

XX Andrews BM, Cantor CR, Edwards CA, Fry KE, Turin LM;

XX WPI; 1999-152755/13.

XX Determination of DNA sequence preference of a DNA-binding molecule -

XX based on inhibition of binding of protein to oligonucleotide

XX sequence attached to test sequence

XX Claim 3; Columns 311-312; 270pp; English.

XX Sequences X17001 to X17600 represent specifically claimed target test  
XX sequences that are used in the method of the invention of determining the  
XX DNA sequence preference of a DNA-binding molecule. The method comprises:  
XX (i) adding a test molecule and a DNA-binding protein to a mixture of  
XX duplex DNA test oligonucleotides, each of the test oligonucleotides  
XX having a test sequence adjacent to a screening sequence, where the  
XX screening sequence binds to the DNA-binding protein with a binding  
XX affinity that is independent of the DNA sequence of the test sequence,  
XX and where the mixture of duplex DNA test oligonucleotides includes  
XX several test sequences; (ii) incubating the test molecule, the mixture of  
XX duplex DNA test oligonucleotides and the DNA-binding protein for a time  
XX sufficient to permit binding of the test molecule to test sequences in  
XX the duplex DNA; (iii) separating unbound test oligonucleotides from test  
XX oligonucleotides bound to binding protein; (iv) amplifying the unbound  
XX test oligonucleotides; (v) repeating steps (ii) to (iv); (vi) isolating  
XX the amplified test oligonucleotides; and (vii) sequencing the isolated  
XX test oligonucleotides. Test sequences X17001-X17481 and X17600 correspond  
XX to promoter targets for human genes and test sequences X17482-X17599  
XX correspond to promoter targets for viral genes.

XX Sequence 50 BP; 26 A; 5 C; 8 G; 11 T; 0 other;

alignment\_scores:

Quality: 28.00 Length: 8  
Ratio: 4.667 Gaps: 0

Percent Similarity: 75.000 Percent Identity: 75.000

alignment\_block:

US-09-277-074-10 x X17415/rev ..

Align seg 1/1 to reverse of: X17415 from: 1 to: 50

2 IlePheGlySerLeuAlaPheLeu 9

|||||  
50 ATCTTGTAGCACACAATTTTA 27

seq\_name: /SIDS6/gcgdata/geneseq/geneseqn/NA1993.DAT.Q36814

seq\_documentation\_block:

ID Q36814 standard; DNA; 55 BP.

XX

AC Q36814;

XX 22-JUN-1993 (first entry)

DE Oligomer SM 81 used in construction of SSP polypeptides.

XX Heptad; plants; custom tailored storage proteins; in vivo;

XX expression; ss.

XX Synthetic.

XX WO9303160-A.

XX 18-FEB-1993.

XX 07-AUG-1992; 92WO-US06412.

XX 09-AUG-1991; 91US-0743006.

XX (DUPO ) DU PONT DE NEMOURS & CO E I.

XX Falco SC, Keeler SJ, Rice JA;

XX WPI; 1993-076317/09.

XX Synthetic polypeptide(s) contg. specified heptad units -

XX expressed in vivo in plants to serve as custom-tailored storage

XX proteins with specified aminoacid content

XX Disclosure; Page 107; 176pp; English.

XX The sequence represents the DNA sequence encoding a synthetic  
XX heptad polypeptide. The synthetic polypeptide can be expressed in  
XX vivo in plants to serve as a synthetic seed storage protein which  
XX can be custom-tailored for specific end-user requirements. The DNA  
XX encoding the heptad may be used to transform plants to increase the  
XX content of partic. amino acids such as lysine or methionine in seeds  
XX or leaves. See also Q36810-28, Q37265-301.

XX Sequence 55 BP; 20 A; 5 C; 23 G; 7 T; 0 other;

alignment\_scores:

Quality: 28.00 Length: 8  
Ratio: 3.900 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 62.500

alignment\_block:

US-09-277-074-10 x Q36814/rev ..

Align seg 1/1 to reverse of: Q36814 from: 1 to: 55

2 IlePheGlySerLeuAlaPheLeu 9

|||||  
36 ATCTTCTTCCATCGCCTTCATC 13

seq\_name: /SIDS6/gcgdata/geneseq/geneseqn/NA1993.DAT.Q36815

seq\_documentation\_block:

ID Q36815 standard; DNA; 55 BP.

```

XX AC Q36815;
XX DT 22-JUN-1993 (first entry)
XX DE Oligomer SM 80 used in construction of SSP polypeptides.
XX KW Heptad; plants; custom tailored storage proteins; in vivo;
XX KW expression; ss.
XX OS Synthetic.
XX PN WO9303160-A.
XX PD 18-FEB-1993.
XX XX 07-AUG-1992; 92WO-US06412.
XX PF 09-AUG-1991; 91US-0743006.
XX PR (DUPO ) DU PONT DE NEMOURS & CO E I.
XX PA Falco SC, Keeler SJ, Rice JA;
XX PI WPI; 1993-076517/09.
XX DR Synthetic polypeptide(s) contg. specified heptad units -
XX XX expressed in vivo in plants to serve as custom-tailored storage
XX PT proteins with specified aminoacid content
XX PT Disclosure; Page 108; 176pp; English.
XX PS The sequence represents the DNA sequence encoding a synthetic
XX CC heptad polypeptide. The synthetic polypeptide can be expressed in
XX CC vivo in plants to serve as a synthetic seed storage protein which
XX CC can be custom-tailored for specific end-user requirements. The DNA
XX CC encoding the heptad may be used to transform plants to increase the
XX CC content of partic. amino acids such as lysine or methionine in seeds
XX CC or leaves. See also Q36810-28, Q37265-301.
XX XX
XX SQ Sequence 55 BP; 8 A; 22 C; 4 G; 21 T; 0 other;

alignment_scores:
  Quality: 28.00 Length: 8
  Ratio: 3.500 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 62.500

alignment_block:
US-09-277-074-10 x Q36815 ..
Align seg 1/1 to: Q36815 from: 1 to: 55
2 IlePheGlySerLeuAlaPheLeu 9
|||||:|||||:|||||:
24 ATCTTCTCTTCATCGCCTTCATC 47

seq_name: /SIDS6/gcgdata/geneseq/geneseqn/NA1995.DAT:Q94972
seq_documentation_block:
ID Q94972 standard; DNA; 55 BP.
XX AC Q94972;
XX DT 16-JUL-1996 (first entry)
XX XX Oligonucleotide SM 81.
XX DE Lysine; synthetic storage protein; SSP; vector; PSK6;
XX KW dihydrodipicolinic acid synthase; corn; maize; Zea mays;
XX KW soybean; Glycine max; transgenic plant; essential amino acid; ss.
XX OS Synthetic.

XX FH Key Location/Qualifiers
XX FT misc_feature 1..55
XX FT /*tag= a
XX FT /standard_name= SM 81
XX FT CDS 2..46
XX FT /*tag= b
XX PN WO9515392-A1.
XX XX 08-JUN-1995.
XX PD 21-NOV-1994; 94WO-US13190.
XX PF 17-JUN-1994; 94US-0261661.
XX PR 30-NOV-1993; 93US-0160117.
XX XX (DUPO ) DU PONT DE NEMOURS & CO E I.
XX PA Falco SC, Keeler SJ, Rice JA;
XX PI WPI; 1995-215272/28.
XX DR P-PSDB; R78236.
XX XX
XX PT New chimeric gene providing increased lysine content in plant seeds
XX PT - contains di:hydro:di:picolinic acid synthase gene coupled to
XX PT chloroplast transport sequence and seed specific promoter, also new
XX PT plants of improved nutritional value.
XX XX
XX PS Example 8; Page 75; 180pp; English.
XX XX
XX CC Oligonucleotides SM81 (Q94972) and SM80 (Q94973) were annealed and
XX CC ligated into vector pSK5. The oligonucleotides code for a base
XX CC peptide (R78236), (SSP5)2, and include an Eari site that allows
XX CC insertion of oligonucleotides coding for 1 or more heptad repeats,
XX CC and a unique Asp718 site for use in transfer of synthetic genes into
XX CC plant vectors. The resulting vector, pSK6, can be used to construct
XX CC synthetic storage protein (SSP) chimeric genes (see also Q94972-Q95005)
XX CC for expression in the seeds of transformed plants, e.g. soybean and
XX CC corn.
XX XX
XX SQ Sequence 55 BP; 20 A; 5 C; 23 G; 7 T; 0 other;

alignment_scores:
  Quality: 28.00 Length: 8
  Ratio: 3.500 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 62.500

alignment_block:
US-09-277-074-10 x Q94972/rev ..
Align seg 1/1 to reverse of: Q94972 from: 1 to: 55
2 IlePheGlySerLeuAlaPheLeu 9
|||||:|||||:|||||:
36 ATCTTCTCTTCATCGCCTTCATC 13

seq_name: /SIDS6/gcgdata/geneseq/geneseqn/NA1995.DAT:Q94973
seq_documentation_block:
ID Q94973 standard; DNA; 55 BP.
XX AC Q94973;
XX DT 15-JUL-1996 (first entry)
XX XX Oligonucleotide SM 80.
XX DE Lysine; synthetic storage protein; SSP; vector; pSK6;
XX KW dihydrodipicolinic acid synthase; corn; maize; Zea mays;
XX KW soybean; Glycine max; transgenic plant; essential amino acid; ss.
XX OS Synthetic.

```

```

OS Synthetic.
XX Key Location/Qualifiers
FH misc_feature 1..55
FT /tag= a
FT /standard_name= SM 80
PN WO9515392-Al.
XX
XX 08-JUN-1995.
XX
XX 21-NOV-1994; 94WO-US13190.
XX
XX 17-JUN-1994; 94US-0261661.
XX 30-NOV-1993; 93US-0160117.
XX
XX (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
XX Falco SC, Keeler SJ, Rice JA;
XX WPI; 1995-215272/28.
XX
XX New chimeric gene providing increased lysine content in plant seeds
PT - contains di:hydro:di:picolinic acid synthase gene coupled to
PT chloroplast transport sequence and seed specific promoter, also new
PT plants of improved nutritional value.
XX
XX Example 8; Page 75; 180pp; English.
XX
XX Oligonucleotides SM81 (Q94972) and SM80 (Q94973) were annealed and
CC ligated into vector pSK5. The oligonucleotides code for a base
CC peptide (R78236), (SP5)2, and include an EarI site that allows
CC insertion of oligonucleotides coding for 1 or more heptad repeats,
CC and a unique Asp718 site for use in transfer of synthetic genes into
CC plant vectors. The resulting vector, pSK5, can be used to construct
CC synthetic storage protein (SSP) chimeric genes (see also Q94972-Q95005)
CC for expression in the seeds of transformed plants, e.g. soybean and
CC corn.
XX
XX Sequence 55 BP; 8'A; 22 C; 4 G; 21 T; 0 other;
SQ

```

```

alignment_scores:
  Quality: 28.00 Length: 8
  Ratio: 3.500 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 62.500

```

```
alignment_block:
US-09-277-074-10 x Q94973 ..
```

```
Align seg 1/1 to: Q94973 from: 1 to: 55
```

```

2 IlePheGlySerLeuAlaPheLeu 9
|||||:|||||:|||||:|
24 ATCTTCTCTCCATCGCCTTCATC 47

```

```
seq_name: /SID56/gcgdata/geneseq/geneseqn/NA1999.DAT:V99505
```

```
seq_documentation_block:
ID V99505 standard; DNA; 55 BP.
```

```
XX V99505;
```

```
XX 29-MAR-1999 (first entry)
```

```
XX Oligonucleotide SM81 used as base gene in vector pSK5.
```

```
XX Lysine; transgenic plant; seed storage protein; vector; pSK5;
KW ds.
```

```
XX Synthetic.
```

```
XX Key Location/Qualifiers
```

```

FT misc_feature 1..4
FT /tag= a
FT /note= "5' single stranded overhang"
FT 55
FT misc_feature
FT /tag= b
FT /note= "5' overhang on complementary strand of
sequence 5'-AATT-3'."
XX
XX WO9842831-A2.
XX
XX 01-OCT-1998;
XX
XX 27-MAR-1998; 98WO-US06051.
XX
XX 27-MAR-1997; 97US-0824627.
XX
XX (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
XX Epelbaum SU, Falco SC, McDevitt RE;
XX WPI; 1999-045139/04.
XX
XX Nucleic acids and chimeric genes for increasing seed lysine content
PT - comprise sequence encoding all or part of lysine ketoglutarate
PT reductase, useful to improve nutritional quality of seeds from
PT transformed plants
XX
XX Example 21; Page 101; 231pp; English.
XX
XX This synthetic double-stranded oligonucleotide comprises a 'base
CC gene' encoding a 14-amino acid peptide. It has been inserted into
CC vector pSK5. The base gene includes an internal EarI site that
CC provides a unique site for subsequent insertion of oligonucleotides
CC encoding one or more heptad repeats. It also includes a 3' Asp718
CC site for use in transfer of gene sequences into plant vectors.
CC Chimeric genes for lysine-rich synthetic seed storage proteins
CC suitable for expression in the seeds of plants (see V99513-18,
CC V99527-32, V99539-41) are inserted into the unique EarI site. The
CC invention provides methods for improving the nutritional quality of
CC seeds from transgenic plants.
XX
XX Sequence 55 BP; 20 A; 5 C; 23 G; 7 T; 0 other;
SQ

```

```

alignment_scores:
  Quality: 28.00 Length: 8
  Ratio: 3.500 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 62.500

```

```
alignment_block:
US-09-277-074-10 x V99505/rev ..
```

```
Align seg 1/1 to reverse of: V99505 from: 1 to: 55
```

```

2 IlePheGlySerLeuAlaPheLeu 9
|||||:|||||:|||||:|
36 ATCTTCTCTCCATCGCCTTCATC 13

```

```
seq_name: /SID56/gcgdata/geneseq/geneseqn/NA1994.DAT:Q63720
```

```
seq_documentation_block:
ID Q63720 standard; DNA; 56 BP.
```

```
XX Q63720;
```

```
XX 25-OCT-1994 (first entry)
```

```
XX Synthetic 2S seed storage protein gene fragment.
```

```
XX Nutrition; non-conserved DNA; assembly; crude; degenerate;
KW oligonucleotides; ss.
```

```
XX Synthetic.
```

```
OS
```

XX WO9410315-A.  
XX  
XX 11-MAY-1994.  
XX  
XX 22-OCT-1993; 93WO-US10090.  
XX  
XX 23-OCT-1992; 92US-0965664.  
XX  
XX (PION-) PIONEER HI-BRED INT INC.  
XX  
XX Ballo B;  
XX  
XX WPI; 1994-167470/20.  
XX  
XX Improving the content of an amino acid in a seed storage protein  
XX to enhance nutritional value - by replacing non-conserved DNA  
XX with DNA encoding the amino acid  
XX  
XX Disclosure; Fig 2; 33pp; English.  
XX  
XX The sequence is that of a fragment of the 2S seed storage gene  
XX which encodes a protein with increased lysine content. The gene  
XX is synthesised from a set of partial DNA sequences capable of being  
XX assembled in a complementary overlapping relationship to provide the  
XX complete DNA. This method does not require a subcloning step and allows  
XX simple, one day assembly of large gene regions. Rapid inclusion of  
XX degenerate oligonucleotide regions can be performed and chimeric  
XX genes can be assembled without introducing mutagenic restriction  
XX sites.  
XX See also Q63710-22.  
XX  
XX Sequence 56 BP; 10 A; 12 C; 13 G; 21 T; 0 other;

alignment\_scores:  
Quality: 28.00 Length: 6  
Ratio: 4.667 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 83.333

alignment\_block:

US-09-277-074-10 x Q63720 ..

Align seg 1/1 to: Q63720 from: 1 to: 56

1 LysilePheGlySerLeu 6  
||||:|||||  
3 AAGGTGTTGGCAGTCTT 20

seq\_name: /SID56/gcgdata/geneseq/geneseq/NAL1997.DAT:T88905

seq\_documentation\_block:

ID T88905 standard; DNA; 28 BP.

XX AC T88905;

XX DT 20-APR-1998 (first entry)

XX DE Human eRF3 cDNA amplifying antisense primer.

XX KW Protein phosphatase 2A; PP2A; release factor; eRF1; binding; cancer;  
XX modulator; protein synthesis; PP2A catalytic subunit; tryptic; PP2Ac.  
XX KW eRF3; PCR primer; ss.

XX OS Synthetic.

XX OS Homo sapiens.

XX PN WO9737037-A1.

XX PD 09-OCT-1997.

XX PF 17-MAR-1997; 97WO-EP01330.

XX

PR 20-DEC-1996; 96GB-0026470.  
PR 29-MAR-1996; 96GB-0006707.  
XX  
XX (NOVS ) NOVARTIS AG.  
XX  
XX Hemmings BA;  
XX  
XX WPI; 1997-503119/46.  
XX  
XX Identifying modulators of protein expression - by screening for  
XX agents which affect the interaction between protein phosphatase 2A  
XX and release factor eRF1.  
XX  
XX Disclosure; Page 10; 27pp; English.

XX  
XX This primer is used for the PCR amplification of the cDNA encoding human  
XX eRF3. The cDNA can be amplified from the human breast carcinoma cell line  
XX T47D total cDNA. Human eRF1 binds to eRF3 and it directly interacts with  
XX protein phosphatase 2A catalytic subunit (PP2Ac). A method for  
XX identifying modulators of protein expression comprises screening for  
XX agents which affect the interaction between PP2Ac and eRF1. The method  
XX comprises incubating eRF1 and PP2Ac or their fragments with the compound  
XX to be screened. Any modulation in the interaction between eRF1 and PP2Ac  
XX can be detected by means of a solid phase binding assay or assessment of  
XX reporter gene expression. The modulators identified can be used for the  
XX regulation of intracellular signalling and protein synthesis. Modulators  
XX of aberrant upregulation of protein synthesis and associated cellular  
XX proliferation can be used for treating diseases associated with cellular  
XX proliferation, particularly cancer.

XX Sequence 28 BP; 5 A; 6 C; 6 G; 11 T; 0 other;

alignment\_scores:

Quality: 27.00 Length: 7  
Ratio: 3.857 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 71.429

alignment\_block:

US-09-277-074-10 x T88905 ..

Align seg 1/1 to: T88905 from: 1 to: 28

3 PheGlySerLeuAlaPheLeu 9

:::|||||:|||||

1 TATGGATCCTTAGTCTTCTC 21

---







```
; APPLICANT: Cantor, Charles R.
; APPLICANT: Andrews, Beth M.
; APPLICANT: Turin, Lisa M.
; APPLICANT: Fry, Kirk E.
; TITLE OF INVENTION: Sequence-Directed DNA Binding
; TITLE OF INVENTION: Molecules, Compositions and Methods
; NUMBER OF SEQUENCES: 641
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genelabs Technologies, Inc.
; STREET: 505 Penobscot Drive
; CITY: Redwood City
; STATE: CA
; COUNTRY: USA
; ZIP: 94063
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: US/08/171,389
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/123,936
; FILING DATE: 17-SEP-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/996,783
; FILING DATE: 23-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/723,618
; FILING DATE: 27-JUN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/081,070
; FILING DATE: 22-JUN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Fabian, Gary R.
; REGISTRATION NUMBER: 33,875
; REFERENCE/DOCKET NUMBER: 4600-0175/G19P3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-0880
; TELEFAX: (415) 324-0960
; INFORMATION FOR SEQ ID NO: 415:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 50 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: Human aldolase B (ALDOB) gene
; US-08-171-389-415

alignment_scores:
  Quality: 28.00      Length: 8
  Ratio: 4.667       Gaps: 0
  Percent Similarity: 75.000      Percent Identity: 75.000

alignment_block:
  US-09-277-074-10 x US-08-171-389-415/rev ..
  Align seg 1/1 to reverse of: US-08-171-389-415 from: 1 to: 50

      2 ilepHeGlySerLeuAlaPheLeu 9
      |||||||||||
      50 ATCTTTGGTAGCACACATTTTTA 27

seq_name: /cgn2_6/ptodata/1/lna/5B_COMB.seq:US-08-123-936-415
seq_documentation_block:
; Sequence 415, Application US/08123936
; Patent No. 5869241
; GENERAL INFORMATION:
; APPLICANT: Cantor, Charles R.
```

```
; Patent No. 5726014
; GENERAL INFORMATION:
; APPLICANT: Edwards, Cynthia A.
; APPLICANT: Cantor, Charles R.
; APPLICANT: Turin, Lisa M.
; APPLICANT: Andrews, Beth M.
; TITLE OF INVENTION: Screening Assay for the Detection of
; TITLE OF INVENTION: DNA-Binding Molecules
; NUMBER OF SEQUENCES: 640
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genelabs Technologies, Inc.
; STREET: 505 Penobscot Drive
; CITY: Redwood City
; STATE: CA
; COUNTRY: USA
; ZIP: 94063
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: US/08/123,936
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/996,783
; FILING DATE: 23-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/723,618
; FILING DATE: 27-JUN-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Fabian, Gary R.
; REGISTRATION NUMBER: 33,875
; REFERENCE/DOCKET NUMBER: 4600-0075.32/G19P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-0880
; TELEFAX: (415) 324-0960
; INFORMATION FOR SEQ ID NO: 415:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 50 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: Human aldolase B (ALDOB) gene
; US-08-123-936-415

alignment_scores:
  Quality: 28.00      Length: 8
  Ratio: 4.667       Gaps: 0
  Percent Similarity: 75.000      Percent Identity: 75.000

alignment_block:
  US-09-277-074-10 x US-08-123-936-415/rev ..
  Align seg 1/1 to reverse of: US-08-123-936-415 from: 1 to: 50

      2 ilepHeGlySerLeuAlaPheLeu 9
      |||||||||||
      50 ATCTTTGGTAGCACACATTTTTA 27

seq_name: /cgn2_6/ptodata/1/lna/5C_COMB.seq:US-08-475-228A-415
seq_documentation_block:
; Sequence 415, Application US/08475228A
; Patent No. 5869241
; GENERAL INFORMATION:
; APPLICANT: Edwards, Cynthia A.
; APPLICANT: Cantor, Charles R.
```

APPLICANT: Andrews, Beth M.  
APPLICANT: Turin, Lisa M.  
APPLICANT: Fry, Kirk E.  
TITLE OF INVENTION: Sequence-Directed DNA Binding  
TITLE OF INVENTION: Molecules, Compositions and Methods  
NUMBER OF SEQUENCES: 664  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genelabs Technologies, Inc.  
STREET: 505 Penobscot Drive  
CITY: Redwood City  
STATE: CA  
COUNTRY: USA  
ZIP: 94063  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/475,228A  
FILING DATE: 06-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/123,936  
FILING DATE: 17-SEP-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/996,783  
FILING DATE: 23-DEC-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/723,618  
FILING DATE: 27-JUN-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/081,070  
FILING DATE: 22-JUN-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Stratford, Carol A.  
REGISTRATION NUMBER: 34,444  
REFERENCE/DOCKET NUMBER: 4600-0175.21/G19P3D2  
TELEPHONE: (415) 324-0880  
TELEFAX: (415) 324-0960  
INFORMATION FOR SEQ ID NO: 415:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 50 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
INDIVIDUAL ISOLATE: Human aldolase B (ALDOB) gene  
US-08-475-228A-415

alignment\_scores:  
Quality: 28.00 Length: 8  
Ratio: 4.667 Gaps: 0  
Percent Similarity: 75.000 Percent Identity: 75.000

alignment\_block:  
US-09-277-074-10 x US-08-475-228A-415/rev ..  
Align seg 1/1 to reverse of: US-08-475-228A-415 from: 1 to: 50

2 IlePheGlySerLeuAlaPheLeu 9  
|||||  
50 ATCTTTGGTAGCACACAATTTTA 27

seq\_name: /cgn2\_6/ptodata/1/lna/6\_COMB.seq:US-08-482-080A-415

seq\_documentation\_block:  
Sequence 415, Application US/08482080A  
Patent No. 6010849  
GENERAL INFORMATION:

APPLICANT: Edwards, Cynthia A.  
APPLICANT: Cantor, Charles R.  
APPLICANT: Andrews, Beth M.  
APPLICANT: Turin, Lisa M.  
APPLICANT: Fry, Kirk E.  
TITLE OF INVENTION: Sequence-Directed DNA Binding  
TITLE OF INVENTION: Molecules, Compositions and Methods  
NUMBER OF SEQUENCES: 664  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genelabs Technologies, Inc.  
STREET: 505 Penobscot Drive  
CITY: Redwood City  
STATE: CA  
COUNTRY: USA  
ZIP: 94063  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/482,080A  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/171,389  
FILING DATE: 20-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/123,936  
FILING DATE: 17-SEP-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/996,783  
FILING DATE: 23-DEC-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/723,618  
FILING DATE: 27-JUN-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/081,070  
FILING DATE: 22-JUN-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Brady, John F.  
REGISTRATION NUMBER: 39,118  
REFERENCE/DOCKET NUMBER: 4600-0175.20/G19P3D1  
TELEPHONE: (650) 324-0880  
TELEFAX: (650) 324-0960  
INFORMATION FOR SEQ ID NO: 415:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 50 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
INDIVIDUAL ISOLATE: Human aldolase B (ALDOB) gene  
US-08-482-080A-415

alignment\_scores:  
Quality: 28.00 Length: 8  
Ratio: 4.667 Gaps: 0  
Percent Similarity: 75.000 Percent Identity: 75.000

alignment\_block:  
US-09-277-074-10 x US-08-482-080A-415/rev ..  
Align seg 1/1 to reverse of: US-08-482-080A-415 from: 1 to: 50

2 IlePheGlySerLeuAlaPheLeu 9  
|||||  
50 ATCTTTGGTAGCACACAATTTTA 27

seq\_name: /cgn2\_6/ptodata/1/lna/PCTUS\_COMB.seq:PCT-US93-12388-415

```
seq_documentation_block:
; Sequence 415, Application PC/TUS9312388
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Sequence-Directed DNA Binding
; TITLE OF INVENTION: Molecules, Compositions and Methods
; NUMBER OF SEQUENCES: 641
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genelabs Technologies, Inc.
; STREET: 505 Penobscot Drive
; CITY: Redwood City
; STATE: CA
; COUNTRY: USA
; ZIP: 94063
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/12388
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/123,936
; FILING DATE: 17-SEP-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/996,783
; FILING DATE: 23-DEC-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Fabian, Gary R.
; REGISTRATION NUMBER: 33,875
; REFERENCE/DOCKET NUMBER: 4600-0175.41/G19PCT2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-0880
; TELEFAX: (415) 324-0960
; INFORMATION FOR SEQ ID NO: 415:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 50 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: Human aldolase B (ALDOB) gene
PCT-US93-12388-415

alignment_scores:
  Quality: 28.00 Length: 8
  Ratio: 4.667 Gaps: 0
  Percent Similarity: 75.000 Percent Identity: 75.000

alignment_block:
US-09-277-074-10 x PCT-US93-12388-415/rev ..

Align seg 1/1 to reverse of: PCT-US93-12388-415 from: 1 to: 50.

2 IlePheGlySerLeuAlaPheLeu 9
|||||
50 ATCTTGTCACACACAAATTTTA 27

seq_name: /cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-182-175A-13

seq_documentation_block:
; Sequence 13, Application US/08182175A
; Patent No. 5559223
; GENERAL INFORMATION:
; APPLICANT: Saverio Carl Falco
; APPLICANT: Sharon J. Keeler
; APPLICANT: Janet A. Rice
; TITLE OF INVENTION: Synthetic Storage Proteins with Defined Structure Containin
; NUMBER OF SEQUENCES: 113
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: E.I. du Pont de Nemours and Company
; STREET: 1007 Market Street
; CITY: Wilmington
```

```

; TITLE OF INVENTION: Synthetic Storage Proteins with Defined Structure Containin
; NUMBER OF SEQUENCES: 113
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: E.I. du Pont de Nemours and Company
; STREET: 1007 Market Street
; CITY: Wilmington
; STATE: Delaware
; COUNTRY: USA
; ZIP: 19898
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: Macintosh
; OPERATING SYSTEM: Macintosh System, 6.0
; SOFTWARE: Microsoft Word, 4.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/182.175A
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/743,006
; FILING DATE: 9 August 1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Linda Axamethy Floyd
; REGISTRATION NUMBER: 33,692
; REFERENCE/DOCKET NUMBER: BB-1031
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (302) 992-4929
; TELEFAX: (302) 892-7949
; TELEX: 835420
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 55 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..55
; OTHER INFORMATION: /product= "synthetic oligonucleotide"
; OTHER INFORMATION: /standard_name= "SM 81"
US-08-182-175A-13

alignment_scores:
  Quality: 28.00 Length: 8
  Ratio: 3.500 Gaps: 0
  Percent Similarity: 100.000 Percent Identity: 62.500

alignment_block:
US-09-277-074-10 x US-08-182-175A-13/rev ..

Align seg 1/1 to reverse of: US-08-182-175A-13 from: 1 to: 55

2 IlePheGlySerLeuAlaPheLeu 9
|||||
36 ATCTTCTTCATCCGCTCATFC 13

seq_name: /cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-182-175A-14

seq_documentation_block:
; Sequence 17, Application US/08182175A
; Patent No. 5559223
; GENERAL INFORMATION:
; APPLICANT: Saverio Carl Falco
; APPLICANT: Sharon J. Keeler
; APPLICANT: Janet A. Rice
; TITLE OF INVENTION: Synthetic Storage Proteins with Defined Structure Containin
; NUMBER OF SEQUENCES: 113
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: E.I. du Pont de Nemours and Company
; STREET: 1007 Market Street
; CITY: Wilmington
```

```
; STATE: Delaware
; COUNTRY: USA
; ZIP: 19898
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: Macintosh
; OPERATING SYSTEM: Macintosh System, 6.0
; SOFTWARE: Microsoft Word, 4.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/182,175A
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/743,006
; FILING DATE: 9 August 1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Linda Axamethy Floyd
; REGISTRATION NUMBER: 33,692
; REFERENCE/DOCKET NUMBER: BB-1031
; TELEPHONE: (302) 992-4929
; TELEFAX: (302) 892-7949
; TELEX: 835420
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 55 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..55
; OTHER INFORMATION: /product= "synthetic oligonucleotide"
; OTHER INFORMATION: /standard_name= "SM 80"
; US-08-182-175A-14

alignment_scores:
  Quality: 28.00      Length: 8
  Ratio: 3.500       Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 62.500

alignment_block:
  US-09-277-074-10 x US-08-182-175A-14 ..
  Align seg 1/1 to: US-08-182-175A-14 from: 1 to: 55

  2 IlePheGlySerLeuAlaPheLeu 9
  |||||:|||||:|||||:
  24 ATCTTCTCTTCATCGCCTTCATC 47

seq_name: /cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-474-633A-20

seq_documentation_block:
; Sequence 20, Application US/08474633A
; Patent No. 5773691
; GENERAL INFORMATION:
; APPLICANT: E. I. DU PONT DE NEMOURS AND
; TITLE OF INVENTION: CHIMERIC GENES AND
; TITLE OF INVENTION: METHODS FOR INCREASING
; TITLE OF INVENTION: INCREASING THE LYSINE
; TITLE OF INVENTION: AND THREONINE CONTENT
; TITLE OF INVENTION: OF THE SEEDS OF PLANTS
; NUMBER OF SEQUENCES: 107
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: E. I. DU PONT DE NEMOURS
; STREET: 1007 MARKET STREET
; CITY: WILMINGTON
; STATE: DELAWARE
; COUNTRY: U.S.A.
```

```
; ZIP: 19898
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: MICROSOFT WORD VERSION 2.0C
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/474,633A
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: BARBARA C. SIEGELL
; REGISTRATION NUMBER: 30,684
; REFERENCE/DOCKET NUMBER: BB-1037-C
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 302-992-4931
; TELEFAX: 302-773-0164
; TELEX: 835420
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 55 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..55
; OTHER INFORMATION: /product= "synthetic
; OTHER INFORMATION: oligonucleotide"
; OTHER INFORMATION: /standard_name= "SM
; OTHER INFORMATION: 81"
; US-08-474-633A-20

alignment_scores:
  Quality: 28.00      Length: 8
  Ratio: 3.500       Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 62.500

alignment_block:
  US-09-277-074-10 x US-08-474-633A-20/rev ..
  Align seg 1/1 to reverse of: US-08-474-633A-20 from: 1 to: 55

  2 IlePheGlySerLeuAlaPheLeu 9
  |||||:|||||:|||||:
  36 ATCTTCTCTTCATCGCCTTCATC 13

seq_name: /cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-474-633A-21

seq_documentation_block:
; Sequence 21, Application US/08474633A
; Patent No. 5773691
; GENERAL INFORMATION:
; APPLICANT: E. I. DU PONT DE NEMOURS AND
; TITLE OF INVENTION: CHIMERIC GENES AND
; TITLE OF INVENTION: METHODS FOR INCREASING
; TITLE OF INVENTION: INCREASING THE LYSINE
; TITLE OF INVENTION: AND THREONINE CONTENT
; TITLE OF INVENTION: OF THE SEEDS OF PLANTS
; NUMBER OF SEQUENCES: 107
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: E. I. DU PONT DE NEMOURS
; STREET: 1007 MARKET STREET
; CITY: WILMINGTON
; STATE: DELAWARE
; COUNTRY: U.S.A.
; ZIP: 19898
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
```



```
; TELEPHONE: (302) 992-4929
; TELEFAX: (302) 892-7949
; TELEX: 835420
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 55 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 1..55
; OTHER INFORMATION: /product= "synthetic oligonucleotide"
; OTHER INFORMATION: /standard_name= "SM 80"
PCT-US92-06412-14
```

```
alignment_scores:
  Quality: 28.00      Length: 8
  Ratio: 3.500       Gaps: 0
Percent Similarity: 100.000 Percent Identity: 62.500
```

```
alignment_block:
US-09-277-074-10 x PCT-US92-06412-14 ..
Align seg 1/1 to: PCT-US92-06412-14 from: 1 to: 55
```

```
2 IlePheGlySerLeuAlaPheLeu 9
|||||:|||||:|||||:
24 ATCTCTCTCCATCGCCTTCATC 47
```

```
seq_name: /cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-682-517-6
```

```
seq_documentation_block:
; Sequence 6, Application US/08682517
; Patent No. 5874267
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Expression of surface layer proteins
; NUMBER OF SEQUENCES: 25
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/682,517
; FILING DATE:
; CLASSIFICATION:
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 44 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-682-517-6
```

```
alignment_scores:
  Quality: 27.00      Length: 9
  Ratio: 3.857       Gaps: 0
Percent Similarity: 77.778 Percent Identity: 66.667
```

```
alignment_block:
US-09-277-074-10 x US-08-682-517-6 ..
Align seg 1/1 to: US-08-682-517-6 from: 1 to: 44
```

```
1 LysIlePheGlySerLeuAlaPheLeu 9
||| :|||||:|||||
2 AAATATTACGGAGCTTTAATTG 28
```

```
seq_name: /cgn2_6/ptodata/1/ina/6_COMB.seq:US-08-985-162-873
```

```
seq_documentation_block:
; Sequence 873, Application US/08985162
; Patent No. 6057156
; GENERAL INFORMATION:
; APPLICANT: Akhtar, Saghir
; APPLICANT: Fell, Patricia
; APPLICANT: McSwiggen, James
; TITLE OF INVENTION: ENZYMIC NUCLEIC ACID TREATMENT
; TITLE OF INVENTION: OF DISEASES OR CONDITIONS RELATED
; TITLE OF INVENTION: TO LEVELS OF EPIDERMAL GROWTH
; TITLE OF INVENTION: FACTOR RECEPTORS
; NUMBER OF SEQUENCES: 1877
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071-2066
```

```
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: FASTSEQ for Windows 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/985,162
; FILING DATE: 04 December 1997
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/036,476
; FILING DATE: 31 January 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard J.
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 230/107
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 873:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 27 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; OTHER INFORMATION: The letter "N" stands for the stem
; OTHER INFORMATION: II region of a HH ribozyme.
US-08-985-162-873
```

```
alignment_scores:
  Quality: 26.00      Length: 8
  Ratio: 3.714       Gaps: 0
Percent Similarity: 87.500 Percent Identity: 75.000
```

```
alignment_block:
US-09-277-074-10 x US-08-985-162-873/rev ..
Align seg 1/1 to reverse of: US-08-985-162-873 from: 1 to: 27
```

```
2 IlePheGlySerLeuAlaPheLeu 9
|||||:|||||:|||||
24 ATCTTTCTCATCAGCCTTCTTA 1
```

```
seq_name: /cgn2_6/ptodata/1/ina/6_COMB.seq:US-08-985-162-1637
seq_documentation_block:
```

Thu Nov 16 15:46:32 2000

```

; Sequence 1637, Application US/08985162
; Patent No. 6057156
; GENERAL INFORMATION:
; APPLICANT: Akhtar, Saghir
; APPLICANT: Fell, Patricia
; APPLICANT: McSwigen, James
; TITLE OF INVENTION: ENZYMATIC NUCLEIC ACID TREATMENT
; TITLE OF INVENTION: OF DISEASES OR CONDITIONS RELATED
; TITLE OF INVENTION: TO LEVELS OF EPIDERMAL GROWTH
; TITLE OF INVENTION: FACTOR RECEPTORS
; NUMBER OF SEQUENCES: 1877
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; STREET: Suite 4700
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071-2066
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: Storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: FastSeq for Windows 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/985,162
; FILING DATE: 04 December 1997
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/036,476
; FILING DATE: 31 January 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard J.
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 230/107
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 1637:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 27 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; OTHER INFORMATION: The letter "N" stands for the stem
; OTHER INFORMATION: II region of a HH ribozyme.
US-08-985-162-1637

```

```

alignment_scores:
    Quality: 26.00      Length: 9
    Ratio: 3.250        Gaps: 0
    Percent Similarity: 88.889    Percent Identity: 55.556

```

```

alignment_block:
US-09-277-074-10 x US-08-985-162-1637/rev ..
Align seg 1/1 to reverse of: US-08-985-162-1637 from: 1 to: 27

```

```

1 LysllepheGlySerLeuAlaphLeu 9
|:::|||||:::|:::|
27 AAAGTATTTCATCATCAGGATTTTG 1

```



OM of: US-09-277-074-10 to: EST: \* out\_format : pfs  
 Date: Nov 15, 2000 10:42 PM  
 About: Results were produced by the GenCore software, version 4.5,  
 Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:  
 -MODE=frame+p2n.model -DEV=xlp  
 -Q=/cgn2\_1/USPRO\_pool/US09277074/runat\_14112000\_120416\_23066/app\_query.fasta\_1.67  
 -DB=EST -OEMT=fastap -SUFFIX=lim60.rst -GAPOP=12.000  
 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000  
 -GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500  
 -GAPOP=6.000 -GAPEXT=7.000 -XGAPOP=10.000 -XGAPEXT=0.500  
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blomsum62  
 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR\_SCORE=pcpt  
 -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFM=pfs  
 -NORM=ext -MINLEN=0 -MAXLEN=60 -USER=US09277074 -CGN1\_1\_1780  
 -NCPU=6 -ICPU=3 -LONGLOG -NO\_XLPXY -WAIT -THREADS=1

Search information block:  
 Query: US-09-277-074-10  
 Query length: 9  
 Database: EST:  
 Database sequences: 7189864  
 Database length: -1203564053  
 Search time (sec): 810.150000

score_list:	Sequence	Strd Orig	Zscore	EScore	Len	Documentation
gb_est21:AW250812	-	29.00	122.57	732.45	47	AW250812 2822529.3prime NIH_MGC
gb_est21:AW250812	-	29.00	120.61	941.42	58	AI360316 qy8605.x1 NCI_CGAP_Brn
gb_est7:AA81957	-	27.00	115.37	1.8e+03	44	AA81957 vx30402.r1 StrataGene
gb_est36:CO1120	-	27.00	113.29	2.4e+03	55	CO1120 HUMGS0007790 Human adult
gb_gss19:BO1268	-	26.00	110.99	3.6e+03	51	BO1268 CSRL-129e10-u CSRL flow s
gb_est8:AI149384	+	25.00	110.82	3.3e+03	31	AI149384 cs86e08.x1 Soares pregn
gb_gss19:BO2304	+	25.00	106.18	6.0e+03	51	BO2304 CSRL-151C12-u CSRL flow s
gb_est7:AA862705	-	25.00	106.00	6.1e+03	52	AA862705 oh40c11.s1 NCI_CGAP_G4
gb_est36:CO121421	-	25.00	105.65	6.4e+03	54	CO121421 HUMGS0009789 Human adult
gb_est10:AI365545	-	25.00	104.99	7.0e+03	58	AI365545 qz0812.x1 NCI_CGAP_G11
gb_est19:AW0708632	+	25.00	104.67	7.3e+03	60	AW0708632 c8a05ne.f1 Neurospora c
gb_gss19:BO2009	+	25.00	104.67	7.3e+03	60	BO2009 CSRL-145F5-u CSRL flow s
gb_est36:CO2292	+	24.00	105.52	6.5e+03	36	CO2292 HUMGS0006658 Human adult
gb_est36:CO121784	+	24.00	102.65	9.4e+03	49	DI1784 HUMH01F06 Liver HepG2 ce
gb_est7:AA933036	+	24.00	102.46	9.7e+03	50	AA933036 oo75e10.s1 NCI_CGAP_K10
gb_est11:AI608417	+	24.00	102.46	9.7e+03	50	AI608417 vk10d06.x1 Knowles Solt
gb_est6:AA713563	+	24.00	101.92	1.0e+04	53	AA713563 nv80f02.s1 NCI_CGAP_Br4
gb_est13:AI904252	+	24.00	101.75	1.1e+04	54	AI904252 PM-BT046-160195-265_2 B
gb_gss19:BO3265	+	24.00	101.57	1.1e+04	55	BO3265 CSRL-173H6-u CSRL flow s
gb_est5:AA647709	+	24.00	101.08	1.2e+04	58	AA647709 vq77h05.s1 Knowles Solt
gb_est36:DI19109	+	24.00	100.76	1.2e+04	60	DI19109 MUSGS01319 Mouse 3'-direc
gb_est34:BE275929	+	23.00	100.87	1.2e+04	39	BE275929 601121563F1 NIH_MGC_20
gb_est36:DI19557	+	23.00	99.96	1.3e+04	43	DI19557 MUSGS00957 Mouse 3'-direc
gb_gss1:AO025319	+	23.00	99.54	1.4e+04	45	AO025319 EP(X)0346 Drosophila me
gb_est10:AI362388	+	23.00	99.33	1.4e+04	46	AI362388 qu5h11.x1 NCI_CGAP_Gas
gb_gss1:AO074099	+	23.00	99.13	1.5e+04	47	AO074099 EP(2)1240 Drosophila me
gb_est4:AA422196	+	23.00	98.74	1.6e+04	49	AA422196 vl16b11.s1 Knowles Solt
gb_est8:AI032978	+	23.00	98.74	1.6e+04	49	AI032978 ex22b01.s1 Soares fetal
gb_gss1:AO073639	+	23.00	98.74	1.6e+04	49	AO073639 EP(2)2502-5prime Drosop
gb_est19:AV564476	+	23.00	98.55	1.6e+04	50	AV564476 AV564476 Arabidopsis th
gb_est11:AV590009	+	23.00	98.19	1.7e+04	52	AV590009 t195a09.x1 NCI_CGAP_Gol
gb_est21:AW272799	+	23.00	98.19	1.7e+04	52	AW272799 xu24b04.x1 NCI_CGAP_Col
gb_est38:N30377	+	23.00	98.19	1.7e+04	52	N30377 yw96a12.s1 Soares placent
gb_est14:AO014031	+	23.00	98.01	1.7e+04	53	AO014031 AU014031 Schizosacchar
gb_est14:AO014274	+	23.00	98.01	1.7e+04	53	AO014274 AU014274 Schizosacchar
gb_est1:AA129201	+	23.00	97.67	1.8e+04	55	AA129201 zn36c06.r1 StrataGene e
gb_est9:AI289974	+	23.00	97.67	1.8e+04	55	AI289974 qy86d10.x1 NCI_CGAP_Ut2
gb_est21:AW245663	+	23.00	97.67	1.8e+04	55	AW245663 hq71d11.x1 NCI_CGAP_HN1
gb_est25:AW872765	+	23.00	97.50	1.8e+04	56	AW872765 hq71d11.x1 NCI_CGAP_HN1
gb_est36:BE536354	+	23.00	97.50	1.8e+04	56	BE536354 601062553F1 NIH_MGC_10
gb_gss21:AG022747	+	23.00	97.34	1.9e+04	57	AG022747 Oryza sativa DNA, 3' fl

gb\_est1:AA124905 - 23.00 97.17 1.9e+04 58 | AA124905 mp73h05.r1 Soares\_th  
 gb\_est6:AA722147 - 23.00 97.17 1.9e+04 58 | AA722147 zh19g09.s1 Soares\_pi  
 gb\_gss19:BO1102 + 23.00 97.17 1.9e+04 58 | BO1102 CSRL-125f10-u CSRL flo  
 gb\_est20:AW130006 + 23.00 97.01 1.9e+04 59 | AW130006 xf26f09.x1 NCI\_CGAP\_

seq\_name: gb\_est21:AW250812

seq\_documentation\_block:  
 LOCUS AW250812 47 bp mRNA EST 07-JAN-2000  
 DEFINITION 2822529.3prime NIH\_MGC\_7 Homo sapiens cDNA clone IMAGE:2822529 3',  
 mRNA sequence.  
 ACCESSION AW250812  
 VERSION AW250812.1 GI:6593805  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 47)  
 AUTHORS NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Other\_ESTS: 2822529.5prime  
 Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 Tissue Procurement: DCTD/DRP cDNA Library Arrayed by: Ling  
 Hong/Rubin Laboratory cDNA Library Arrayed by: The I.M.A.G.E.  
 Consortium (LLNL) DNA Sequencing by: Berkeley MGC sequencing  
 project clone distribution: MGC clone distribution information can  
 be found through the I.M.A.G.E. Consortium/LLNL at:  
 www.bio.llnl.gov/bbrp/image/html Base Calling / Quality  
 Scores: PHRED from University of Washington Genome Center  
 Trimming: crossmatch from University of Washington Genome Center  
 PHRAP suite. Poly-T Identification: patmatch.pl from Berkeley  
 Drosophila Genome Project. University of Washington Genome Center:  
 http://www.genome.washington.edu Low Quality Sequence: 10  
 contiguous PHRED high quality bases following vector sequence. Very  
 Low Quality Sequence: Trace file contained 47 contiguous distinct  
 peaks following vector sequence. Polyadenylation: Based upon the  
 presence of a XhoI site followed by a run of 14 or more T residues  
 at the beginning of the sequence, this cDNA insert was  
 polyadenylated.  
 Plate: L1Cw9 row: K column: 10  
 High quality sequence stop: 10.  
 Location/Qualifiers  
 1..47  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_image="2822529"  
 /clone\_lib="NIH\_MGC\_7"  
 /tissue\_type="small cell carcinoma"  
 /cell\_line="DHI03"  
 /lab\_host="DHI08 (phage-resistant)"  
 /notes="Organ: lung; Vector: pOTB7; Site: 1: XhoI; Site: 2:  
 EORI; cDNA made by oligo-dT priming. Directionally  
 cloned into EORI/XhoI sites using the following 5',  
 adaptor: GCCACGAG(G). Size-selected >500bp for average  
 insert size 1.8kb. Library constructed by Ling Hong in  
 the laboratory of Gerald M. Rubin (University of  
 California, Berkeley) using ZAP-cDNA synthesis kit  
 (Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT  
 ORIGIN

alignment\_scores:  
 Quality: 29.00 Length: 8  
 Ratio: 4.143 Gaps: 0  
 Percent Similarity: 87.500 Percent Identity: 62.500

alignment\_block:  
 US-09-277-074-10 x AW250812/rev



27 TTTGGGCTCTAGCTTTTATC 7

seq\_name: gb\_est36:c01120

seq\_documentation\_block:

LOCUS C01120 55 bp mRNA 23-JUL-1996  
 DEFINITION HUMGS0007790 Human adult (K.Okubo) Homo sapiens cDNA, mRNA sequence.

ACCESSION C01120

VERSION C01120.1 GI:1433350

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 55)

BodyMap: human gene expression database

Unpublished (1995)

JOURNAL

COMMENT

Contact: Okubo,K.

Institute for Molecular and Cellular Biol

Osaka University

1-3,Yamada-oka, Suita, Osaka Pref. 565, Japan

Tel: 06-877-5111(ex.3315)

Email: kousaku@imcb.osaka-u.ac.jp

Human Gene Signature, 3'-directed cDNA sequence. We are not

submitting the same cDNA sequence redundantly to DBJ since 1993.

For the abundance information of clones with this sequence in this

library and as well as in other 3'-directed libraries, see

http://www.imcb.osaka-u.ac.jp/bodymap/. The sequences of the clones

represented by this GS sequences is also found there.

FEATURES

source

1..55

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_lib="Human adult (K.Okubo)"

/dev\_stage="adult"

BASE COUNT 21 a 10 c 9 g 15 t

ORIGIN

alignment\_scores:

Quality: 27.00

Ratio: 3.857

Percent Similarity: 87.500

Length: 8

Gaps: 0

Percent Identity: 62.500

alignment\_block:

US-09-277-074-10 x C01120/rev ..

Align seg 1/1 to reverse of: C01120 from: 1 to: 55

1 LyslePheGlySerLeuAlaPhe 8

|||||:|||||:|||||

35 AAGTATTCTCAAGTCGTATT 12

seq\_name: gb\_gss19:B01268

seq\_documentation\_block:

LOCUS B01268 51 bp DNA 13-JUL-1996  
 DEFINITION CSRL-129e10-u CSRL flow sorted Chromosome 11 specific cosmid Homo sapiens genomic clone CSRL-129e10, DNA sequence.

ACCESSION B01268

VERSION B01268.1 GI:1410546

KEYWORDS GSS.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 51)

Evans,G.A., Burbee,D., Davies,C., Hahner,L., Oliver,T., Gilbert,M.,

Jones,D., Ward,T., Gillilan,E., Schagmann,J., Probst,S., Harris

J., DeFord,J., McFarland,J., Burzinski,K., Khan,M., Kupfer,K. and

Garner,H.R.

REFERENCE

AUTHORS

J., DeFord,J., McFarland,J., Burzinski,K., Khan,M., Kupfer,K. and

Garner,H.R.

TITLE

JOURNAL

COMMENT

Genomic Sequence Sampled Map of Chromosome 11

Unpublished (1996)

Contact: Evans GA, Shane Probst

McDermott Center for Human Growth and Development

University of Texas Southwestern Medical Center At Dallas

5323 Harry Hines Blvd, Dallas TX 75235-8591

Tel: 214-648-1600

Fax: 214-648-1666

Email: gevas@utsw.swmed.edu, shane@mcdermott.swmed.edu

Seq primer: 17

Class: cosmid ends

High quality sequence stop: 51.

FEATURES

source

1..51

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_lib="CSRL-129e10"

/cosmid="CSRL flow sorted Chromosome 11 specific

cosmid"

/sex="female"

/cell\_type="chimeric hamster somatic cell hybrid"

/note="Vector: scos-1; Human Chromosome 11 specific cosmid

library prepared from flow sorted human Chromosome 11

derived from Chinese Hamster Ovary (CHO) monochromosomal

somatic cell hybrid, J1"

BASE COUNT 22 a 11 c 6 g 11 t 1 others

ORIGIN

alignment\_scores:

Quality: 26.00

Ratio: 4.333

Percent Similarity: 100.000

Length: 6

Gaps: 0

Percent Identity: 83.333

alignment\_block:

US-09-277-074-10 x B01268/rev ..

Align seg 1/1 to reverse of: B01268 from: 1 to: 51

3 PheGlySerLeuAlaPhe 8

|||||:|||||:|||||

33 TTTGGATCGCTGGGTTT 16

seq\_name: gb\_est8:A1149384

seq\_documentation\_block:

LOCUS A1149384

DEFINITION qc86e08.x1 Soares pregnant uterus\_NbHPU Homo sapiens CDNA clone

IMAGE:1721126 3' similar to TR:Q14218 Q14218 ENDOTHELIAL CELL

PROTEIN C/APC RECEPTOR PRECURSOR. ;, mRNA sequence.

ACCESSION A1149384

VERSION A1149384.1 GI:3577853

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 31)

NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov

This clone is available royalty-free through LBNL ; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Trace considered overall poor quality

Seq primer: -40m3 fwd. ET from Amersham

High quality sequence stop: 1.

Location/Qualifiers

1..31

/organism="Homo sapiens"

```

/db_xref="taxon:9606"
/clone="IMAGE:I721126"
/clone_lib="Soares_pregnant_uterus_NbHPU"
/sex="female"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: uterus; Vector: pT73-Pac; Site: 1; Not I;
Site2: Eco RI; 1st strand cDNA was primed with a Not I -
oligo(dT) primer [5',
AACTGGAAGATTCGGCGCCCTTTTTTTTTTTTTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization. Library
constructed by M. Fatima Bonaldo."
BASE COUNT      3 a      8 c      7 g      13 t
ORIGIN

alignment_scores:
  Quality:      25.00      Length:      7
  Ratio:        3.571      Gaps:      0
Percent Similarity: 100.000 Percent Identity: 57.143

alignment_block:
US-09-277-074-10 x A1149384 ..
Align seg 1/1 to: A1149384 from: 1 to: 31

1 LysilePheGlySerLeuAla 7
:::|||||
7 CCTGTTGGCTCCCTTCG 27

seq_name: gb_gss19:B02304

seq_documentation_block:
LOCUS      B02304      51 bp      DNA      GSS      13-JUL-1996
DEFINITION CSRL-151C12-u CSRL flow sorted Chromosome 11 specific cosmid Homo
sapientis genomic clone CSRL-151C12, DNA sequence.
ACCESSION      B02304
VERSION      B02304.1 GI:1411582
KEYWORDS      GSS.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS      Evans G.A., Burbee,D., Davies,C., Hahner,L., Oliver,T., Gilbert,M.,
Jones,D., Ward,T., Gillilan,E., Schagemann,J., Probst,S., Harris
J., DeFord,J., McFarland,J., Burzinski,K., Khan,M., Kupfer,K. and
Garner,H.R.
TITLE      Genomic Sequence Sampled Map of Chromosome 11
JOURNAL      Unpublished (1996)
COMMENT      Contact: Evans GA, Shane Probst
McDermott Center for Human Growth and Development
University of Texas Southwestern Medical Center At Dallas
5323 Harry Hines Blvd, Dallas TX 75235-8591
Tel: 214-648-1600
Fax: 214-648-1666
Email: geavans@utsw.swmed.edu, shane@mcdermott.swmed.edu
Seq primer: T7
Class: cosmid ends
High quality sequence stop: 51.
Location/Qualifiers
1..51
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CSRL-151C12"
/clone_lib="CSRL flow sorted Chromosome 11 specific
cosmid"
/sex="female"
/note="Vector: sCos-1; Human Chromosome 11 specific cosmid

library prepared from flow sorted human Chromosome 11
derived from Chinese Hamster Ovary (CHO) monochromosomal
somatic cell hybrid, J1"
BASE COUNT      21 a      8 c      7 g      14 t      1 others
ORIGIN

alignment_scores:
  Quality:      25.00      Length:      7
  Ratio:        4.167      Gaps:      0
Percent Similarity: 85.714 Percent Identity: 57.143

alignment_block:
US-09-277-074-10 x B02304/rev ..
Align seg 1/1 to reverse of: B02304 from: 1 to: 51

2 IlePheGlySerLeuAlaPhe 8
:::|||||
47 TTGTCGGGTCATATTTT 27

seq_name: gb_est7:AA862705

seq_documentation_block:
LOCUS      AA862705      52 bp      mRNA      EST      13-MAY-1998
DEFINITION ch40c11.s1 NCI_CGAP_GC4 Homo sapiens cDNA clone IMAGE:1469108 3',
similar to TR:P92496 P92496 NADH DEHYDROGENASE SUBUNIT 2 ;, mRNA
sequence.
ACCESSION      AA862705
VERSION      AA862705.1 GI:2955184
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS      NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE      National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL      Unpublished (1997)
COMMENT      Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution Information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bbrp/image/image.html

Trace considered overall poor quality
Insert Length: 708 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 1.
Location/Qualifiers
1..52
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1469108"
/clone_lib="NCI_CGAP_GC4"
/tissue_type="pooled germ cell tumors"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; 1st strand cDNA was prepared from 3 pooled
germ cell tumors, and was then primed with a Not I -
oligo(dT) primer. Double-stranded cDNA was ligated to Eco
RI adaptors (Pharmacia), digested with Not I and cloned
into the Not I and Eco RI sites of the modified pT73
vector. Library is normalized. Library was constructed by
Bento Soares and M. Fatima Bonaldo."
BASE COUNT      32 a      0 c      13 g      7 t

```

```

ORIGIN
alignment_scores:
  Quality: 25.00      Length: 8
  Ratio: 3.571       Gaps: 0
Percent Similarity: 87.500 Percent Identity: 62.500

alignment_block:
US-09-277-074-10 x AA862705/rev ..
Align seg 1/1 to reverse of: AA862705 from: 1 to: 52
2 IlePheGlySerLeuAlaPheLeu 9
:::|||||:|||||:|||||
27 TTATTTCTCTCTCTATTCCTT 4

seq_name: gb_est36:C21421

seq_documentation_block:
LOCUS C21421 54 bp mRNA EST 23-OCT-1996
DEFINITION HMG50009789 Human adult (K.Okubo) Homo sapiens cDNA 3', mRNA
sequence.
ACCESSION C21421
VERSION C21421.1 GI:1622531
KEYWORDS EST.
SOURCE human.
ORGANISM
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
  1 (bases 1 to 54)
  Tissue Procurement: Ash Alizadeh, John Byrd, M.D., Mike Grever,
  M.D., Louis M. Staudt, M.D., Ph.D.
  CDNA Library Preparation: M. Bento Soares, Ph.D.
  CDNA Library Arrayed by: Greg Lennon, Ph.D.
  DNA Sequencing by: Washington University Genome Sequencing Center
  Clone distribution: NCI-CGAP clone distribution information can be
  found through the I.M.A.G.E. Consortium/LLNL at:
  www-bio.llnl.gov/bbrp/image/image.html
FEATURES
  source
    1..54
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone_lib="Human adult (K.Okubo)"
    /dev_stage="adult"
    /tissue_type="B-cell, chronic lymphocytic leukemia"
    /lab_host="DH10B"
    /note="Vector: pT73D-Pac (Pharmacia) with a modified
    polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
    was primed with a Not I - oligo(dT) primer (5'
    TGTTACCAATGCTGAAGTGGAGCGCGCATGCTTTTCTTTTCTTTTCTTTT
    T 3'); double-stranded cDNA was ligated to Eco RI
    adaptors (Pharmacia), digested with Not I and cloned into
    the Not I and Eco RI sites of the modified pT73 vector.
    Library is normalized, and was constructed by Bento
    Soares and M.Fatima Bonaldo."
BASE COUNT 17 a 17 c 9 g 21 t
ORIGIN

alignment_scores:
  Quality: 25.00      Length: 6
  Ratio: 5.000       Gaps: 0
Percent Similarity: 83.333 Percent Identity: 83.333

alignment_block:
US-09-277-074-10 x C21421/rev ..
Align seg 1/1 to reverse of: C21421 from: 1 to: 54
3 PheGlySerLeuAlaPhe 8
|||||
26 TTGGGCACCTTAGCATTT 9

seq_name: gb_est10:A1365545
seq_documentation_block:
LOCUS AW708632 60 bp mRNA EST 25-APR-2000
DEFINITION c8a05ne.fl Neurospora crassa evening cDNA library Neurospora crassa
cDNA clone c8a05ne 5', mRNA sequence.
ACCESSION AW708632

```

---

```

LOCUS A1365545 58 bp mRNA EST 07-JAN-1999
DEFINITION g208q12.x1 NCI-CGAP_CLL1 Homo sapiens cDNA clone IMAGE:2020966 3'
similar to SW:PRPL_HUMAN P10162 SALIVARY PROLINE-RICH PROTEIN PO
; contains THR.B3 THR repetitive element ;, mRNA sequence.
ACCESSION A1365545
VERSION A1365545.1 GI:4125234
KEYWORDS EST.
SOURCE human.
ORGANISM
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
  1 (bases 1 to 58)
  NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
  Tumor Gene Index
  Unpublished (1997)
  Contact: Robert Strausberg, Ph.D.
  Tel: (301) 496-1550
  Email: Robert.Strausberg@nih.gov
  Tissue Procurement: Ash Alizadeh, John Byrd, M.D., Mike Grever,
  M.D., Louis M. Staudt, M.D., Ph.D.
  CDNA Library Preparation: M. Bento Soares, Ph.D.
  CDNA Library Arrayed by: Greg Lennon, Ph.D.
  DNA Sequencing by: Washington University Genome Sequencing Center
  Clone distribution: NCI-CGAP clone distribution information can be
  found through the I.M.A.G.E. Consortium/LLNL at:
  www-bio.llnl.gov/bbrp/image/image.html
FEATURES
  source
    1..58
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone_lib="NCI-CGAP_CLL1"
    /tissue_type="B-cell, chronic lymphocytic leukemia"
    /lab_host="DH10B"
    /note="Vector: pT73D-Pac (Pharmacia) with a modified
    polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
    was primed with a Not I - oligo(dT) primer (5'
    TGTTACCAATGCTGAAGTGGAGCGCGCATGCTTTTCTTTTCTTTTCTTTT
    T 3'); double-stranded cDNA was ligated to Eco RI
    adaptors (Pharmacia), digested with Not I and cloned into
    the Not I and Eco RI sites of the modified pT73 vector.
    Library is normalized, and was constructed by Bento
    Soares and M.Fatima Bonaldo."
BASE COUNT 20 a 16 c 19 g 3 t
ORIGIN

alignment_scores:
  Quality: 25.00      Length: 8
  Ratio: 4.167       Gaps: 0
Percent Similarity: 75.000 Percent Identity: 50.000

alignment_block:
US-09-277-074-10 x A1365545/rev ..
Align seg 1/1 to reverse of: A1365545 from: 1 to: 58
1 LysIlePheGlySerLeuAlaPhe 8
:::|||||:|||||
28 CGGTGTTCGGACCCCTCCCTTTT 5

seq_name: gb_est24:AW708632
seq_documentation_block:
LOCUS AW708632 60 bp mRNA EST 25-APR-2000
DEFINITION c8a05ne.fl Neurospora crassa evening cDNA library Neurospora crassa
cDNA clone c8a05ne 5', mRNA sequence.
ACCESSION AW708632

```

```

VERSION      AW708632.1  GI:7597679
KEYWORDS     EST.
SOURCE       Neurospora crassa.
ORGANISM     Eukaryota; Fungi; Ascomycota; Sordariales; Sordariaceae;
             Neurospora.
REFERENCE    1 (bases 1 to 60)
AUTHORS      Zhu, H., Lai, H., Kupfer, D., Dunlap, J.C. and Roe, B.A.
TITLE        Two Neurospora crassa EST Databases
JOURNAL      Unpublished (1998)
COMMENT      Other ESTs: c8a05ne.r1
             Contact: Bruce A. Roe, University of Oklahoma, broe@ou.edu
             Department of Chemistry and Biochemistry
             Advanced Center for Genome Technology, University of Oklahoma
             620 Parrington Oval, Norman, OK 73019, USA
             Tel: 405 325 4912
             Fax: 405 325 7762
             Email: broe@ou.edu
             We anticipate the future release of the cDNA clones to the Fungal
             Genetics Stock Center
             Seq primer: Universal Forward Primer
             High quality sequence stop: 50.
FEATURES     source
             1..60
             /organism="Neurospora crassa"
             /strain="Strain 30-7 (bd: A)"
             /db_xref="taxon:5141"
             /clone="c8a05ne"
             /tissue_type="Neurospora crassa evenng cDNA library"
             /note="Vector: pBluescript SK-; Site_1: XbaI; Site_2:
             EcoRI; See: Bell-Pedersen, D., et al. PNAS 93:13096,1996.
             5' end of cDNA cloned into XbaI site of pBluescript; 3'
             end of cDNA cloned into EcoRI site of pBluescript"
BASE COUNT   15 a 18 c 18 g 9 t
ORIGIN

alignment_scores:
  Quality: 25.00      Length: 7
  Ratio: 4.167       Gaps: 0
  Percent Similarity: 85.714  Percent Identity: 71.429

alignment_block:
  US-09-277-074-10 x AW708632 ..
  Align seg 1/1 to: AW708632 from: 1 to: 60
  1 LysillePheGlySerLeuAla 7
  |||::: |||:::|||||
  32 AAGGTTTCACGGAAGTTTGCC 52
seq_name: gb_gss19:B02009
seq_documentation_block:
  LOCUS      B02009          60 bp      DNA
  DEFINITION  CSRL-145F5-u CSRL flow sorted Chromosome 11 specific cosmid Homo
             sapiens genomic clone CSRL-145F5, DNA sequence.
  ACCESSION  B02009
  VERSION    B02009.1  GI:1411287
  KEYWORDS   GSS.
  SOURCE     human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1 (bases 1 to 60)
AUTHORS      Evans, G.A., Burbee, D., Davies, C., Hahner, L., Oliver, T., Gilbert, M.,
             Jones, D., Ward, T., Gillilan, E., Schagemann, J., Probst, S., Harris
             J., DeFord, J., McFarland, J., Burzinski, K., Khan, M., Kupfer, K. and
             Garner, H.R.
TITLE        Genomic Sequence Sampled Map of Chromosome 11
JOURNAL      Unpublished (1996)

```

---

```

COMMENT      Contact: Evans GA, Shane Probst
             McDermott Center for Human Growth and Development
             University of Texas Southwestern Medical Center At Dallas
             5323 Harry Hines Blvd, Dallas TX 75235-8591
             Tel: 214-648-1600
             Fax: 214-648-1666
             Email: gevangs@utsw.swmed.edu, shane@mcdermott.swmed.edu
             Seq primer: T7
             Class: cosmid ends
             High quality sequence stop: 60.
FEATURES     source
             1..60
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             /clone="CSRL-145F5"
             /clone_lib="CSRL flow sorted Chromosome 11 specific
             cosmid"
             /sex="female"
             /cell_type="chimeric hamster somatic cell hybrid"
             /note="Vector: sCos-1; Human Chromosome 11 specific cosmid
             library prepared from flow sorted human Chromosome 11
             derived from Chinese Hamster Ovary (CHO) monochromosomal
             somatic cell hybrid, J1"
BASE COUNT   25 a 11 c 12 g 11 t 1 others
ORIGIN

alignment_scores:
  Quality: 25.00      Length: 6
  Ratio: 4.167       Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 66.667

alignment_block:
  US-09-277-074-10 x B02009/rev ..
  Align seg 1/1 to reverse of: B02009 from: 1 to: 60
  3 PheGlySerLeuAlaPhe 8
  |||:::|||||:::|||||
  52 TTGGGAGTATTCTTTT 35
seq_name: gb_est36:C02292
seq_documentation_block:
  LOCUS      C02292          36 bp      mRNA
  DEFINITION  HUMGS0006658 Human adult (K.Okubo) Homo sapiens cDNA, mRNA
             sequence.
  ACCESSION  C02292
  VERSION    C02292.1  GI:1434522
  KEYWORDS   EST.
  SOURCE     human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1 (bases 1 to 36)
AUTHORS      Okubo, K.
TITLE        Bodymap; human gene expression database
JOURNAL      Unpublished (1995)
COMMENT      Contact: Okubo, K.
             Institute for Molecular and Cellular Biol
             Osaka University
             1-3,Yamada-oka, Suita, Osaka Pref. 565, Japan
             Tel: 06-877-5111(ex.3315)
             Email: kousaku@imcb.osaka-u.ac.jp
             Human Gene Signature, 3'-directed cDNA sequence. We are not
             submitting the same cDNA sequence redundantly to DBJ since 1993.
             For the abundance information of clones with this sequence in this
             library and as well as in other 3'-directed libraries, see
             http://www.imcb.osaka-u.ac.jp/bodymap/. The sequences of the clones
             represented by this GS sequences is also found there.
FEATURES     source
             1..36
             /organism="Homo sapiens"

```

/db\_xref="taxon:9606"  
 /clone\_lib="Human adult (K.Okubo)"  
 /dev\_stage="adult" 13 a 6 c 4 g 13 t

BASE COUNT  
 ORIGIN

alignment\_scores:  
 Quality: 24.00 Length: 5  
 Ratio: 4.800 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-277-074-10 x C02292 ..

Align seg 1/1 to: C02292 from: 1 to: 36

2 IlePheGlySerLeu 6

|||||

21 ATCTTGGCAGCCTT 35

seq\_name: gb\_est36:D11784

seq\_documentation\_block: 49 bp mRNA EST 02-DEC-1992  
 LOCUS D11784  
 DEFINITION HUM01F06 Liver HepG2 cell line. Homo sapiens cDNA clone hm01f06,  
 mRNA sequence.

ACCESSION D11784

VERSION D11784.1 GI:2155064

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 49)

AUTHORS Okubo,K., Hori,N., Matoba,R., Niiyama,T., Fukushima,A., Kojima,Y.  
 and Matsubara,K.

TITLE Large scale cDNA sequencing for analysis of quantitative and

qualitative aspects of gene expression

Nature Genet. 2, 173-179 (1992)

94258199

COMMENT Contact: Kousaku Okubo, Naohiro Hori, Ryo Matoba, Toshiyuki Niiyama

, Atsushi Fukushima, Yuko Kojima & Kenichi Matsubara

Institute for Molecular and Cellular Biology

Osaka University

1-3 Yamada-oka, Suita, Osaka 565, Japan.

Location/Qualifiers

FEATURES

source

1..49

/organism="Homo sapiens"

/db\_xref="GDB:D0S8330E"

/db\_xref="taxon:9606"

/clone="hm01f06"

/clone\_lib="Liver HepG2 cell line."

/lab\_host="E.coli"

/note="3'-directed regional cDNA library. Cleaved by MboI  
 and transformed into E.coli."

BASE COUNT

ORIGIN

alignment\_scores:

Quality: 24.00 Length: 6

Ratio: 4.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 83.333

alignment\_block:

US-09-277-074-10 x D11784/rev ..

Align seg 1/1 to reverse of: D11784 from: 1 to: 49

4 GlySerLeuAlaPheLeu 9

|||||

41 GGGTCACTGNGCTTTTA 24

seq\_name: gb\_est7:AA933036

seq\_documentation\_block:

LOCUS AA933036

DEFINITION OO75e10.g1 NCI\_CGAP\_Kid5 Homo sapiens cDNA clone IMAGE:1572042 3',  
 similar to SW:E2BD.RABIT P4111 TRANSLATION INITIATION FACTOR  
 EIF-2B DELTA SUBUNIT ;, mRNA sequence.

ACCESSION AA933036

VERSION AA933036.1 GI:3086969

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 50)

AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index

JOURNAL Unpublished (1997)

COMMENT

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert\_Strausberg@nih.gov

Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.

Emmert-Buck, M.D., Ph.D.

cDNA Library Preparation: M. Bento Soares, Ph.D.

cDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 235 Std Error: 0.00

Seq Primer: -40ml3 fwd. ET from Amersham.

FEATURES

source

1..50

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:1572042"

/clone\_lib="NCI\_CGAP\_Kid5"

/tissue\_type="2 pooled tumors (clear cell type)"

/lab\_host="DH10B"

/note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with  
 a modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st  
 strand cDNA was primed with a Not I - oligo(dT) primer [5',  
 AACTGGAGAAATTCGCCGCCCATATTTTTTTTTTTT 3'],  
 double-stranded cDNA was ligated to Eco RI adaptors  
 (Pharmacia), digested with Not I and cloned into the Not I  
 and Eco RI sites of the modified pT7T3 vector. Library  
 went through one round of normalization. Library  
 constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 8 a 9 c 9 g 24 t

ORIGIN

alignment\_scores:

Quality: 24.00 Length: 6

Ratio: 4.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 66.667

alignment\_block:

US-09-277-074-10 x AA933036 ..

Align seg 1/1 to: AA933036 from: 1 to: 50

4 GlySerLeuAlaPheLeu 9

|||||

15 GGGAGTATGCGCATTTATT 32

---